PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/40, C07K 14/18, C12Q 1/70, C07K 16/10, G01N 33/576

A2

(11) International Publication Number: WO 96/13590

(43) International Publication Date: 9 May 1996 (09.05.96)

(21) International Application Number: PCT/EP95/04155

(22) International Filing Date: 23 October 1995 (23.10.95)

(30) Priority Data:
94870166.9 21 October 1994 (21.10.94) EP
(34) Countries for which the regional or international application was filed: AT et al.
95870076.7 28 June 1995 (28.06.95) EP
(34) Countries for which the regional or international application was filed: AT et al.

(71) Applicant *(for all designated States except US)*: INNOGENETICS N.V. [BE/BE]; Industriepark Zwijnaarde 7, Box 4, B-9052 Zwijnaarde (BE).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: MAERTENS, Geert [BE/BE]; Zilversparrenstraat 64, B-8310 Brugge (BE). STUYVER, Lieven [BE/BE]; Holestraat 8, B-2400 Mol (BE).

(74) Common Representative: INNOGENETICS N.V.; Mrs. Ann De Clercq, Industriepark Zwijnaarde 7, Box 4, B-9052 Zwijnaarde (BE).

(81) Designated States: AL, AM, AT, AU, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TT, UA, UG, US, UZ, VN, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG), ARIPO patent (KE, LS, MW, SD, SZ, UG).

Published
Without international search report and to be republished upon receipt of that report.
With a request for rectification under Rule 91.1(f).

(54) Title: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS

(57) Abstract

The present invention relates to new genomic nucleotide sequences and amino acid sequences corresponding to the coding region of these genomes. The invention relates to new HCV types and subtypes sequences which are different from the known HCV types and subtypes sequences. More particularly, the present invention relates to new HCV type 7 sequences, new HCV type 9 sequences, new HCV type 10 and new HCV type 11 sequences. Also, the present invention relates to new HCV type 1 sequences of subtypes 1d, 1e, 1f and 1g; new HCV type 2 sequences of subtypes 2e, 2f, 2g, 2h, 2i, 2k and 2l; new HCV type 3 sequences of subtype 3g, new HCV type 4 sequences of subtypes 4k, 4l and 4m; a process for preparing them, and their use for diagnosis, prophylaxis and therapy. More particularly, the present invention provides new type-specific sequences of the Core, the E1 and the NS5 regions of new HCV types 7, 9, 10 and 11, as well as of new variants (subtypes) of HCV types 1, 2, 3 and 4. These new HCV sequences are useful to diagnose the presence of HCV type 1, and/or type 2, and/or type 3, and/or type 4, and/or type 7, and/or type 9, and/or type 10, and/or type 11 genotypes or serotypes in a biological sample. Moreover, the availability of these new type-specific sequences can increase the overall sensitivity of HCV detection and should also prove to be useful for prophylactic and therapeutic purposes.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS

The invention relates to new sequences of hepatitis C virus (HCV) genotypes and their use as prophylactic, therapeutic and diagnostic agents.

The present invention relates to new genomic nucleotide sequences and amino acid sequences corresponding to the coding region of these genomes. The invention relates to new HCV types and subtypes sequences which are different from the known HCV types and subtypes sequences. More particularly, the present invention relates to new HCV type 7 sequences, new HCV type 9 sequences, new HCV types 10 and new HCV type 11 sequences. Also the present invention relates to new HCV type 1 sequences of subtypes 1d, 1e, 1f and 1g; new HCV type 2 sequences of subtypes 2e, 2f, 2g, 2h, 2i, 2k and 2l; new HCV type 3 sequences of subtype 3g, new HCV type 4 sequences of subtypes 4k, 4l and 4m; a process for preparing them, and their use for diagnosis, prophylaxis and therapy.

The technical problem underlying the present invention is to provide new HCV sequences from untill now unknown HCV types and/or subtypes. More particularly, the present invention provides new type-specific sequences of the Core, the E1 and the NS5 regions of new HCV types 7, 9, 10 and 11, as well as of new variants (subtypes) of HCV types 1, 2, 3 and 4. These new HCV sequences are useful to diagnose the presence of HCV type 1, and/or type 2, and/or type 3, and/or type 4, and/or type 7, and/or type 9, and/or type 10, and/or type 11 genotypes or serotypes in a biological sample. Moreover, the availability of these new type-specific sequences can increase the overall sensitivity of HCV detection and should also prove to be useful for prophylactic and therapeutic purposes.

Hepatitis C viruses (HCV) have been found to be the major cause of non-A, non-B hepatitis. The sequences of cDNA clones covering the complete genome of several prototype isolates have been determined (Kato et al., 1990; Choo et al., 1991; Okamoto et al., 1991; Okamoto et al., 1992). Comparison of these isolates shows that the variability in nucleotide sequences can be used to distinguish at least 2 different genotypes, type 1 (HCV-1 and HCV-J) and type 2 (HC-J6 and HC-J8),

with an average homology of about 68%. Within each type, at least two subtypes exist (e.g. represented by HCV-1 and HCV-J), having an average homology of about 79%. HCV genomes belonging to the same subtype show average homologies of more than 90% (Okamoto et al., 1992). However, the partial nucleotide sequence of the NS5 region of the HCV-T isolates showed at most 67% homology with the previously published sequences, indicating the existence of yet another HCV type (Mori et al., 1992). Parts of the 5' untranslated region (UR), core, NS3, and NS5 regions of this type 3 have been published, further establishing the similar evolutionary distances between the 3 major genotypes and their subtypes (Chan et al., 1992). Type 4 was subsequently discovered (Stuyver et al., 1993b; Simmonds et al., 1993a; Bukh et al., 1993; Stuyver et al., 1994a). As well as type 5 (Stuyver et al., 1993b; Simmonds et al., 1993c; Bukh et al., 1993; Stuyver et al., 1994b), and type 6 HCV groups (Bukh et al., 1993; Simmonds et al., 1993c). An overview of the present state of the art regarding HCV genotypes is given in Table 3. The nomenclature system proposed by the inventors of the present application has now been accepted by scientists worldwide (Simmonds et al., 1994).

The aim of the present invention is to provide new HCV nucleotide and amino acid sequences enabling the detection of HCV infection.

Another aim of the present infection is to provide new nucleotide and amino acid HCV sequences enabling the classification of infected biological fluids into different serological groups.

Another aim of the present invention is to provide new nucleotide and amino acid HCV sequences ameliorating the overall HCV detection rate.

Another aim of the present invention is to provide new HCV sequences, useful for the design of HCV prophylactic or therapeutic vaccine compositions.

Another aim of the present invention is to provide a pharmaceutical composition consisting of antibodies raised against the polypeptides encoded by these new HCV sequences, for therapy or diagnosis.

All the aims of the present invention are met by the following embodiments of the present invention.

The present invention relates more particularly to an HCV polynucleic acid, having a nucleotide sequence which is unique to a heretofore unidentified HCV type or subtype which is different from HCV subtypes 1a, 1b, 1c, 2a, 2b, 2c, 2d, 3a, 3b,

3c, 3d, 3e, 3f, 4a, 4b, 4c, 4d, 4e, 4f, 4g, 4h, 4i, 4j, 5a or 6a, with said HCV subtypes being classified as in Table 3 by comparison of a part of the NS5 gene nucleotide sequence spanning positions 7932 to 8271, with said amino acid numbering being shown in Table 1, and with said polynucleic acid containing at least one nucleotide differing from said known HCV nucleotide sequences, or the complement thereof. The sequence of known HCV isolates may be found in any nucleotide sequence database known in the art (such as for instance the EMBL database).

The present invention thus also relates to a polynucleic acid having a nucleotide sequence which is unique to at least one of HCV subtypes 1d, 1e, 1f, 1g, 2e, 2f, 2g, 2h, 2i, 2k, 2l, 3g, 4k, 4l, 4m, 7a, 7c or 7d, with said HCV subtypes being classified as defined above.

The present invention thus also relates to a polynucleic acid having a nucleotide sequence which is unique to at least one of HCV types 9, 10 or 11, with said HCV types being classified as defined above.

It is to be noted that the nucleotide(s) difference in the polynucleic acids of the invention may involve an amino acid difference in the corresponding amino acid sequences encoded by said polynucleic acids. A composition according to the present invention may contain only polynucleic acid sequences or polynucleic acid sequences mixed with any excipient known in the art of diagnosis, prophylaxis or therapy.

According to a preferred embodiment, the present invention relates to a polynucleic acid encoding an HCV polyprotein comprising in its amino acid sequence at least one of the following amino acid residues:

I15, C38, V44, A49, Q43, P49, Q55, A58, S60 or D60, E68 or V68, H70, A71 or Q71 or N71, D72, H81, H101, D106, S110, L130, I134, E135, L140, S148, T150 or E150, Q153, F155, D157, G160, E165, I169, F181, L186, T190, T192 or I192 or H192, I193, A195, S196, R197 or N197 or K197, Q199 or D199 or H199 or N199, F200 or T200, A208, I213, M216 or S216, N217 or S217 or G217 or K217, T218, I219, A222, Y223, I230, W231 or L231, S232 or H232 or A232, Q233, E235 or L235, F236 or T236, F237, L240 or M240, A242, N244, N249, I250 or K250 or R250, A252 or C252, A254, I255 or V255, D256 or M256, E257, E260 or K260, R261, V268, S272 or R272, I285, G290 or F290, A291, A293 or L293

or W293, T294 or A294, S295 or H295, K296 or E296, Y297 or M297, I299 or Y299, I300, S301, P316, S2646, A2648, G2649, A2650, V2652, Q2653, H2656 or L2656, D2657, F2659, K2663 or Q2663, A2667 or V1667, D2677, L2681, M2686 or Q2686 or E2686, A2692 or K2692, H2697, I2707, L2708 or Y2708, A2709, A2719 or M2719, F2727, T2728 or D2728, E2729, F2730 or Y2730, I2741, I2745, V2746 or E2746 or L2746 or K2746, A2748, S2749 or P2749, R2750, E2751, D2752 or N2752 or S2752 or T2752 or V2752 or I2752 or Q2752, S2753 or D2753 or G2753, D2754, A2755, L2756 or Q2756, R2757,

with said notation being composed of a letter representing the amino acid residue by its one-letter code, and a number representing the amino acid numbering according to Kato et al. (1980), as shown in Table 1,

or a part of said polynucleic acid which is unique to at least one of the HCV subtypes or types as defined in Table 5, and which contains at least one nucleotide differing from known HCV nucleotide sequences, or the complement thereof.

Each of the above-mentioned residues can be found in Figures 2, 4 or 6 showing the new amino acid sequences of the present invention aligned with known sequences of other types or subtypes of HCV for the Core/E1 region.

According to another preferred embodiment, the present invention relates to a polynucleic acid encoding a HCV polyprotein comprising in its amino acid sequence at least one amino acid sequence chosen from the following list:

ARQSDGRSWAQ or ARRSEGRSWAQ as for subtype 1d and 108)	(SEQ ID NO 107
ERRPEGRSWAQ as for subtype 1e	(SEQ ID NO 109)
ARRPEGRSWAQ as for subtype 1f	(SEQ ID NO 110)
DRRTTGKSWGR as for subtype 2k	(SEQ ID NO 111)
DRRATGRSWGR as for subtype 2e	(SEQ ID NO 112)
DRRATGKSWGR as for subtype 2f	(SEQ ID NO 113)
VRQPTGRSWGQ as for type 9	(SEQ ID NO 114)
VRHQTGRTWAQ as for subtype 7a and 7c	(SEQ ID NO 115)
VRQNQGRTWAQ as for subtype 7d	(SEQ ID NO 116)
ARRTEGRSWAQ as for type 10	(SEQ ID NO 117)
VRRTTGRXXXX or VRRTTGRTWAQ as for type 11 and 119)	(SEQ ID NO 118

HEVRNASGVYHV or HEVRNASGVYHL as for subtype 1d (SEQ ID NO 120 and 121)

YEVHSTTDGYHV as for subtype 1f (SEQ ID NO 122)

VEVKNTSQAYMA as for subtype 2e (SEQ ID NO 123)

IQVKNNSHFYMA as for subtype 2f (SEQ ID NO 124)

VQVKNTSTMYMA as for subtype 2g (SEQ ID NO 125)

VQVKNTSHSYMV as for subtype 2h (SEQ ID NO 126)

VQVANRSGSYMV as for subtype 2i (SEQ ID NO 127)

VEIKNTXNTYVL or VEIKNTSNTYVL as for subtype 2k (SEQ ID NO 128 and 129)

INYRNVSGIYYV or INYRNTSGIYHV or INYHNTSGIYHI or TNYRNVSGIYHV as for subtype 4k (SEQ ID NO 130, 131, 132 or 133)

QHYRNVSGIYHV as for subtype 4l (SEQ ID NO 134)

IQVKNASGIYHL as for type 9 (SEQ ID NO 135)

AHYTNKSGLYHL as for subtype 7c (SEQ ID NO 136)

LNYANKSGLYHL as for subtype 7d (SEQ ID NO 137)

LEYRNASGLYMV as for type 10 (SEQ ID NO 138)

IYEMDGMIMHY or IYEMSGMILHA as for subtype 1d (SEQ ID NO 139 and 140)

VYEAKDIILHT as for subtype 1f (SEQ ID NO 141)

VWQLXDAVLHV as for subtype 2e (SEQ ID NO 142)

VWQLRDAVLHV as for subtype 2f (SEQ ID NO 143)

IWQMQGAVLHV as for subtype 2g (SEQ ID NO 144)

VWQLKDAVLHV as for subtype 2h (SEQ ID NO 145)

VWQLEEAVLHV as for subtype 2i (SEQ ID NO 146)

TWQLXXAVLHV as for subtype 2k (SEQ ID NO 147)

VYEADHHILHL or VYEADHHILAL or VFEADHHILHL as for subtupe 4k (SEQ ID NO 148, 149 and 150)

VYESDHHILHL as for subtype 4l (SEQ ID NO 151)

VFEAETMILHL as for type 9 (SEQ ID NO 152)

VYEAETLILHL as for subtype 7c (SEQ ID NO

153)

VYEANGMILHL as for subtype 7d (SEQ ID NO 154)

VYEAGDIILHL as for type 10 (SEQ ID NO 155)

VREDNHLRCWMAL or VRENNSSRCWMAL as for subtype 1d (SEQ ID NO 156 and 157)

IREGNISRCWVPL as for subtype 1f (SEQ ID NO 158)

ENSSGRFHCWIPI as for subtype 2e (SEQ ID NO 159)

ERSGNRTFCWTAV as for subtype 2f (SEQ ID NO 160)

ELQGNKSRCWIPV as for subtype 2g (SEQ ID NO 162)

ERHQNQSRCWIPV as for subtype 2h (SEQ ID NO 163)

EWKDNTSRCWIPV as for subtype 2i (SEQ ID NO 164)

EREGNSSRCWIPV as for subtype 2k (SEQ ID NO 165)

VREGNQSRCWVAL or VRTGNQSRCWVAL or VRVGNQSSCWVAL or VRVGNQSRCWVAL or VKEGNHSRCWVAL as for subtype 4k (SEQ ID NO 166, 167, 168 or 169)

VKTGNTSRCWVAL as for subtype 4l (SEQ ID NO 170)

IKAGNESRCWLPV as for type 9 (SEQ ID NO 171)

VKEGNQSRCWVQA as for subtype 7c (SEQ ID NO 172)

VKXXNLTKCWLSA as for subtype 7d (SEQ ID NO 173)

VRSGNTSRCWIPV as for type 10 (SEQ ID NO 174)

VKNASVPTAA or VKDANVPTAA as for subtype 1d (SEQ ID NO 175 and 176)

ARIANAPIDE as for subtype 1f (SEQ ID NO 177)

VSKPGALTKG as for subtype 2e (SEQ ID NO 178)

VSRPGALTRG as for subtype 2f (SEQ ID NO 179)

VNQPGALTRG as for subtype 2g (SEQ ID NO 180)

VSQPGALTRG as for subtype 2h (SEQ ID NO 181)

VSQPGALTKG as for subtype 2i (SEQ ID NO 182)

VSRPGALTEG as for subtype 2k (SEQ ID NO 183)

APYIGAPLES or APYTAAPLES as for subtype 4k (SEQ ID NO 184 and 185)

APILSAPLMS as for subtype 4l (SEQ ID NO 186)

VPNSSVPIHG as for type 9 (SEQ ID NO 187)

VPNASTPVTG as for subtype 7c (SEQ ID NO 188)

VQNASVSIRG as for subtype 7d (SEQ ID NO 189)
VKSPCAATAS as for type 10 (SEQ ID NO 190)
SPRMHHTTQE or SPRLYHTTQE as for subtype 1d (SEQ ID NO 191 and 192)
TSRRHWTVQD as for subtype 1f (SEQ ID NO 193)
APKRHYFVQE as for subtype 2e (SEQ ID NO 194)
SPQYHTFVQE as for subtype 2f (SEQ ID NO 195)
SPQHHNFSQD as for subtype 2g (SEQ ID NO 196)
SPQHHIFVQD as for subtype 2h (SEQ ID NO 197)
SPEHHHFVQD as for subtype 2k (SEQ ID NO 198)
RPRRHWTTQD or RPRRHWTAQD or QPRRHWTTQD or RPRRHWTTQE as for subtype 4k (SEQ ID NO 199, 200, 201 or 202)
QPRRHWTVQD as for subtype 4l (SEQ ID NO 203)
RPKYHQVTQD as for type 9 (SEQ ID NO 204)
RPRMHQVVQE as for subtype 7c (SEQ ID NO 205)
RPRMYEIAQD as for subtype 7d (SEQ ID NO 206)
RHRQHWTVQD as for type 10 (SEQ ID NO 207)

or a part of said polynucleic acid which is unique to at least one of the HCV subtypes or types as defined Table 5, and which contains at least one nucleotide differing from known HCV nucleotide sequences, or the complement thereof.

Using the 5' non-coding LiPA system (Stuyver et al., 1993) and a new core LiPA system including multiple probes for subtypes 1a, 1b, 1c, 2a, 2b or 2c derived from the core region (Stuyver et al., 1995), samples from the Benelux, Cameroon, France and Vietnam were selected because of their aberrant reactivities (isolates CAM1078, FR2, FR1, VN4, VN12, VN13, NE98). Some samples were, together with many other samples, sequenced as a control for typing. Sequencing results, however, indicated the discovery of new subtypes (isolates BNL1, BNL2, BNL3, FR4, BNL4, BNL5, BNL6, BNL7, BNL8, BNL9, BNL10, BNL11 and BNL12). Nucleotide sequences in the core and E1 regions which have not yet been reported before, were analyzed in the frame of the invention. Genomic sequences of subtype 1d, 1e, 1f, 1g 2e, 2f, 2g, 2h, 2i, 2k, 2l, 3g, 4k, 4l, 4m, 7a, 7c, 7d and types 9, 10 and 11 isolates are reported for the first time in the present invention. The NS5B region was also analyzed.

The term "polynucleic acid" refers to a single- stranded or double-stranded

nucleic acid sequence which may contain at least 5 contiguous nucleotides in common with the complete nucleotide sequence (e.g. at least 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50, 60, 75 or more contiguous nucleotides). A polynucleic acid which is up till about 100 nucleotides in length is often also referred to as an oligonucleotide. A polynucleic acid may consist of deoxyribonucleotides or ribonucleotides, nucleotide analogues or modified nucleotides, or may have been adapted for therapeutic purposes. A polynucleic acid may also comprise a double stranded cDNA clone which can be used for cloning purposes, or for *in vivo* therapy, or prophylaxis.

The oligonucleotides according to the present invention, used as primers or probes may also contain or consist of nucleotide analogous such as phosphorothioates (Matsukura et al., 1987), alkylphosphoriates (Miller et al., 1979) or peptide nucleic acids (Nielsen et al., 1991; Nielsen et al., 1993) or may contain interculating agents (Asseline et al., 1984).

As most other variations or modifications introduced into the original DNA sequences of the invention these variations will neccissitate adaptions with respect to the conditions under which the oligonucleotide should be used to obtain the required specificty and sensitivity. However the eventual results will be essentially the same as those obtained with the unmodified oligonucleotides.

The introduction of these modifications may be advantageous in order to positivily influence characteristics such as hybridization kinetics, reversibility of the hybrid-formation, biological stability of the oligonucleotide molecules, etc.

The polynucleic acids of the invention may be comprised in a composition of any kind. Said composition may be for diagnostic, therapeutic or prophylactic use.

The expression "sequences which are unique to an HCV type or subtype" refers to sequences which are not shared by any other type or subtype of HCV, and can thus be used to uniquely detect that HCV type or subtype. Sequence variability is demonstrated in the present invention between the newly found HCV types and subtypes (see Table 5) and the known HCV types and subtypes (see Table 3), and it is therefore from these regions of sequence variability in particular that type- or subtypes-specific polynucleic acids, oligonucleotides, polypeptides and peptides may be obtained. The term type- or subtypes-specific refers to the fact that a sequence is unique to that HCV type or subtype involved.

The expression "nucleotides corresponding to" refers to nucleotides which are homologous or complementary to an indicated nucleotide sequence or region within a specific HCV sequence.

The term "coding region" corresponds to the region of the HCV genome that encodes the HCV polyprotein. In fact, it comprises the complete genome with the exception of the 5' untranslated region and 3' untranslated region.

The term "HCV polyprotein" refers to the HCV polyprotein of the HCV-J isolate (Kato et al., 1990). The adenine residue at position 330 (Kato et al., 1990) is the first residue of the ATG codon that initiates the long HCV polyprotein of 3010 amino acids in HCV-J and other type 1b isolates, and of 3011 amino acids in HCV-1 and other type 1a isolates, and of 3033 amino acids in type 2 isolates HC-J6 and HC-J8 (Okamoto et al., 1992).

This adenine is designated as position 1 at the nucleic acid level, and this methionine is designated as position 1 at the amino acid level, in the present invention. As type 1a isolates contain 1 extra amino acid in the NS5A region, coding sequences of type 1a and 1b have identical numbering in the Core, E1, NS3, and NS4 region, but will differ in the NS5B region as indicated in Table 1. Type 2 isolates have 4 extra amino acids in the E2 region, and 17 or 18 extra amino acids in the NS5 region compared to type 1 isolates, and will differ in numbering from type 1 isolates in the NS3/4 region and NS5b regions as indicated in Table 1. Similar insertions compared with type 1 (but of a different size) can also be observed in type 3a sequences which affect the numbering of type 3a amino acids accordingly. Other insertions or deletions may be readily observed in type1, 2, 3, 4, 5, 6, 7, 8, 9, 10 and 11 sequences after alignment withknown HCV sequences.

TABLE 1

	Region	Positions described in the present invention*	Positions described for HCV-J (Kato et al., 1990)	Positions described for HCV-1 (Choo et al., 1991)	Positions described for HC-J6, HC-J8 (Okamoto et al., 1992)

Nucleotides	NS5B	8023/8235 7932/8271	8352/8564 8261/8600	8026/8238 7935/8274	8433/8645 8342/8681
		coding region of present invention	330/9359	1/9033	342/9439
Amino Acids	NS5B	2675/2745 2645/2757	2675/2745 2645/2757	2676/2746 2646/2758	2698/2768 2668/2780

Table 1: Comparison of the HCV nucleotide and amino acid numbering system used in the present invention (*) with the numbering used for other prototype isolates. For example, 8352/8564 indicates the region designated by the numbering from nucleotide 8352 to nucleotide 8564 as described by Kato et al. (1990). Since the numbering system of the present invention starts at the polyprotein initiation site, the 329 nucleotides of the 5' untranslated region described by Kato et al. (1990) have to be substracted, and the corresponding region is numbered from nucleotide 8023 ('8352-329') to 8235 ('8564-329').

The term "genotype" as used in the present invention refers to both types and/or subtypes.

The term "HCV type" corresponds to a group of HCV isolates of which the complete genome shows more than 73% preferably more than 74% homology at the nucleic acid level, or of which the NS5 region between nucleotide positions 7932 and 8271 shows more than 75.4% homology at the nucleic acid level, or of which the complete HCV polyprotein shows more than 78% homology at the amino acid level, or of which the NS5 region between amino acids at positions 2645 and 2757 shows more than 80% homology at the amino acid level, to polyproteins of the other isolates of the group, with said numbering beginning at the first ATG codon or first methionine of the long HCV polyprotein of the HCV-J isolate (Kato et al., 1990). Isolates belonging to different types of HCV exhibit homologies, over the complete genome, of less than 74%, preferably less than 73%, at the nucleic acid level and less than 78% at the amino acid level. Isolates belonging to the same type usually

show homologies of about 90 to 99% at the nucleic acid level and 95 to 96% at the amino acid level when belonging to the same subtype, and those belonging to the same type but different subtypes preferably show homologies of about 76% to 82% (more particularly of about 77% to 80%) at the nucleic acid level and 85-86% at the amino acid level.

More preferably the definition of HCV types is concluded from the classification of HCV isolates according to their nucleotide distances calculated as detailed below:

(1) based on phylogenetic analysis of nucleic acid sequences in the NS5B region between nucleotides 7935 and 8274 (Choo et al., 1991) or 8261 and 8600 (Kato et al., 1990) or 8342 and 8681 (Okamoto et al., 1991), isolates belonging to the same HCV type show nucleotide distances of less than 0.34, usually less than 0.33, and more usually of less than 0.32, and isolates belonging to the same subtype show nucleotide distances of less than 0.135, usually of less than 0.13, and more usually of less than 0.125, usually ranging between 0.0003 and 0.1151, and consequently isolates belonging to the same type but different subtypes show nucleotide distances ranging from 0.135 to 0.34, usually ranging from 0.1384 to 0.2977, and more usually ranging from 0.15 to 0.32, and isolates belonging to different HCV types show nucleotide distances greater than 0.34, usually greater that 0.35, and more usually of greater than 0.358, more usually ranging from 0.3581 to 0.6670.

(2) based on phylogenetic analysis of nucleic acid sequences in the core/E1 region between nucleotides 378 and 957, isolates belonging to the same HCV type show nucleotide distances of less than 0.38, usually of less than 0.37, and more usually of less than 0.364, and isolates belonging to the same subtype show nucleotide distances of less than 0.17, usually of less than 0.16, and more usually of less than 0.15, more usually less than 0.135, more usually less than 0.134, and consequently isolates belonging to the same type but different subtypes show nucleotide distances ranging from 0.15 to 0.38, usually ranging from 0.16 to 0.37, and more usually ranging from 0.17 to 0.36, more usually ranging from 0.133 to 0.379, and isolates belonging to different HCV types show nucleotide distances greater than 0.34, 0.35, 0.36, usually more than 0.365, and more usually of greater than 0.37,

Table 2 : Molecular evolutionary distances

Region	Core/E1 579 bp	E1 384 bp	NS5B 340 bp	NS5B 222 bp
Isolates*	0.0017 - 0.1347 (0.0750 ± 0.0245)	0.0026 - 0.2031 (0.0969 ± 0.0289)	0.0003 - 0.1151 (0.0637 ± 0.0229)	0.000 - 0.1323 (0.0607 ± 0.0205)
Subtypes*	0.1330 - 0.3794 (0.2786 ± 0.0363)	0.1645 - 0.4869 (0.3761 ± 0.0433)	0.1384 - 0.2977 (0.2219 ± 0.0341)	0.117 - 0.3538 (0.2391 ± 0.0399)
Types*	0.3479 - 0.6306 (0.4703 ± 0.0525)	0.4309 - 0.9561 (0.6308 ± 0.0928)	0.3581 - 0.6670 (0.4994 ± 0.0495)	0.3457 - 0.7471 (0.5295 ± 0.0627)

Table 2 Figures created by the PHYLIP program DNADIST are expressed as minimum to maximum (average ± standard deviation). Phylogenetic distances for isolates belonging to the same subtype ('isolates'), to different subtypes of the same type ('subtypes'), and to different types ('types') are given.

In a comparative phylogenetic analysis of available sequences, ranges of molecular evolutionary distances for different regions of the genome were calculated, based on 19,781 pairwise comparisons by means of the DNADIST program of the phylogeny inference package PHYLIP version 3.5c (Felsenstein, 1993). The results are shown in Table 2 and indicate that although the majority of distances obtained in each region fit with classification of a certain isolate, only the ranges obtained in the 340bp NS5B-region are non-overlapping and therefore conclusive. However, as was performed in the present invention, it is preferable to obtain sequence information from at least 2 regions before final classification of a given isolate.

Designation of a number to the different types of HCV and HCV nomenclature is based on chronological discovery of the different types. The numbering system used in the present invention might still fluctuate according to international conventions or guidelines. For example, "type 4" might be changed into "type 5" or "type 6". Also the arbitrarily chosen border distances between types and subtypes and isolates may still be subject to change according to international guidelines or

conventions. Therefore types 7a, 8a, 8b, 9a may for example be designated 6b, 6c, 6d, and 6d in the future; and type 10a which shows relatedness with genotype 3 may be denoted 3g instead of 10a.

The term "subtype" corresponds to a group of HCV isolates of which the complete polyprotein shows a homology of more than 90% both at the nucleic acid and amino acid levels, or of which the NS5 region between nucleotide positions 7932 and 8271 shows a homology of more than 90% at the nucleic acid level to the corresponding parts of the genomes of the other isolates of the same group, with said numbering beginning with the adenine residue of the initiation codon of the HCV polyprotein. Isolates belonging to the same type but different subtypes of HCV show homologies of more than 74% at the nucleic acid level and of more than 78% at the amino acid level.

It is to be understood that extremely variable regions such as the E1, E2 and NS4 regions will exhibit lower homologies than the average homology of the complete genome of the polyprotein.

Using these criteria, HCV isolates can be classified into at least 11 types. Several subtypes can clearly be distinguished in types 1, 2, 3, 4 and 7 : 1a, 1b, 1c, 1d, 1e, 1f, 1g, 2a, 2b, 2c, 2d, 2e, 2f, 2g, 2h, 2i, 2k, 2l, 3a, 3b, 3c, 3d, 3f, 3g, 4a, 4b, 4c, 4d, 4e, 4f, 4g, 4h, 4i, 4j, 4k, 4l, 4m, 7a, 7c, and 7d based on homologies of the 5' UR and coding regions. An overview of most of the reported isolates and their proposed classification according to the typing system of the present invention as well as other proposed classifications is presented in Table 3.

Table 3

HCV CLASSIFICATION

	OKA-MOTO	MORI	CHA	NAKAO	PROTOTYPE
1a	I	I	Pt	GI	HCV-1, HCV-H, HC-J1
1b	II	II	KI	GII	HCV-J, HCV-BK, HCV-T, HC-JK1, HC-J4, HCV-CHINA
1c					HC-G9
2a	III	III	K2a	GIII	HC-J6
2b	IV	IV	K2b	GIII	HC-J8

2c					S83, ARG6, ARG8, I10, T983
2d					NE92
3a	V	V	K3	GIV	BR36, BR56, HD10, N2L1, BR33, Ta, E-b1
3b		VI	K3	GIV	HCV-TR, Tb, NE137
3c					NE48
3d					NE274
3e					NE145
3f					NE125
4a					Z4, GB809-4
4b					Z1
4c					GB116, GB358, GB215, Z6, Z7
4d					DK13
4e					GB809-2, CAM600, CAM736
4f					CAM622, CAM627
4g					GB549
4h					GB438
4i					CAR4/1205
4j					CAR1/905
5a				GV	SA3, SA4, SA1, SA7, SA11, BE95
6a					HK1, HK2, HK3, HK4, VN11

Table 3 Overview of the known HCV types and subtypes classified according to the different authors.

The term "complement" refers to a nucleotide sequence which is complementary to an indicated sequence and which is able to hybridize to the indicated sequences.

The composition of the invention can comprise many combinations. By way of example, the composition of the invention can comprise:

- two (or more) nucleic acids from the same region or,
- two nucleic acids (or more), respectively from different regions, for the same isolate or for different isolates,
- or nucleic acids from the same regions and from at least two different regions

(for the same isolate or for different isolates).

The present invention relates particularly to a polynucleic acid as defined above having a sequence selected from any of SEQ ID NO 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39 , 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103 to 105, or a part of said polynucleic acid which is unique to any of the HCV subtypes or types as defined in Table 5, and which contains at least one nucleotide differing from known HCV polynucleic acids, or the complement thereof.

The present invention relates more particularly to a polynucleic acid as defined above, which codes for the 5' UR, the Core/E1, the NS4 or the NS5B region or a part thereof.

More particularly, the present invention relates to a polynucleic acid as defined above which is a cDNA sequence.

Also included within the present invention are sequence variants of the polynucleic acids as selected from any of the nucleotide sequences as given in any of the above given SEQ ID numbers with said sequence variants containing either deletion and/or insertions of one or more nucleotides, especially insertions or deletions of 1 or more codons, mainly at the extremities of oligonucleotides (either 3' or 5'), or substitutions of some non-essential nucleotides (i.e. nucleotides not essential to discriminate between different genotypes of HCV) by others (including modified nucleotides an/or inosine), for example, a type 1 or 2 sequence might be modified into a type 7 sequence by replacing some nucleotides of the type 1 or 2 sequence with type-specific nucleotides of type 7 as shown in for instance Figure 1 and 2.

Particularly preferred variant polynucleic acids of the present invention include also sequences which hybridise under stringent conditions with any of the polynucleic acid sequences of the present invention. Particularly, sequences which show a high degree of homology (similarity) to any of the polynucleic acids of the invention as described above. Particularly sequences which are at least 80%, 85%, 90%, 95% or more homologous to said polynucleic acid sequences of the invention. Preferably said sequences will have less than 20%, 15%, 10%, or 5% variation of the original nucleotides of said polynucleic acid sequence.

Polynucleic acid sequences according to the present invention which are

homologous to the sequences as represented by a SEQ ID NO can be characterized and isolated according to any of the techniques known in the art, such as amplification by means of sequence-specific primers, hybridization with sequence-specific probes under more or less stringent conditions, serological screening methods or via the LiPA typing system.

Other preferred variant polynucleic acids of the present invention include sequences which are redundant as a result of the degeneracy of the genetic code compared any of the above-given polynucleic acids of the present invention. These variant polynucleic acid sequences will thus encode the same amino acid sequence as the polynucleic acids they are derived from.

Also included within the scope of the present invention are 5' non-coding region sequences which can be readily obtained from type 1 subtype 1d, 1e, 1f or 1g isolates; type 2 subtype 2e, 2f, 2g, 2h, 2i, 2k or 2l isolates; type 3 subtype 3g isolates; type 4 subtype 4k, 4l or 4m isolates; type 7 subtype 7a, 7c or 7d isolates, type 9, type 10 or type 11 isolates discribed herein. Such sequences may contain type or subtype-specific motifs which can be employed for type and/or subtype-specific hybridization assays, e.g. such as described by Stuyver et al. (1993).

Polynucleic acid sequences of the genomes indicated above from regions not yet depicted in the present examples, figures and sequence listing can be obtained by any of the techniques known in the art, such as amplification techniques using suitable primers from the sequences of these new genomes given in Figure 1 of the present invention.

The present invention also relates to an oligonucleotide primer comprising part of a polynucleic acid as defined above, with said primer being able to act as a primer for specifically amplifying the nucleic acid of a certian HCV isolate belonging to the genotype from which the primer is derived.

The term "primer" refers to a single stranded DNA oligonucleotide sequence capable of acting as a point of initiation for synthesis of a primer extension product which is complementary to the nucleic acid strand to be copied. The length and the sequence of the primer must be such that they allow to prime the synthesis of the extension products. Preferably the primer is about 5-50 nucleotides. Specific length and sequence will depend on the complexity of the required DNA or RNA targets, as well as on the conditions of primer use such as temperature and ionic strength.

The fact that amplification primers do not have to match exactly with corresponding template sequence to warrant proper amplification is amply documented in the literature (Kwok et al., 1990).

The amplification method used can be either polymerase chain reaction (PCR; Saiki et al., 1988), ligase chain reaction (LCR; Landgren et al., 1988; Wu & Wallace, 1989; Barany, 1991), nucleic acid sequence-based amplification (NASBA; Guatelli et al., 1990; Compton, 1991), transcription-based amplification system (TAS; Kwoh et al., 1989), strand displacement amplification (SDA; Duck, 1990; Walker et al., 1992) or amplification by means of Qß replicase (Lizardi et al., 1988; Lomeli et al., 1989) or any other suitable method to amplify nucleic acid molecules using primer extension. During amplification, the amplified products can be conveniently labelled either using labelled primers or by incorporating labelled nucleotides. Labels may be isotopic (^{32}P, ^{35}S, etc.) or non-isotopic (biotin, digoxigenin, etc.). The amplification reaction is repeated between 20 and 70 times, advantageously between 25 and 45 times.

The present invention also relates to an oligonucleotide probe comprising part of a polynucleic acid as defined above, with said probe being able to act as a hybridization probe for specific detection and/or classification into types and/or subtypes of an HCV nucleic caid containing said nucleotide sequence, with said probe being possibly labelled or attached to a solid substrate.

The term "probe" refers to single stranded sequence-specific oligonucleotides which have a sequence which is complementary to the target sequence of the HCV genotype(s) to be detected.

Preferably, these probes are about 5 to 50 nucleotides long, more preferably from about 10 to 25 nucleotides.

The term "solid support" can refer to any substrate to which an oligonucleotide probe can be coupled, provided that it retains its hybridization characteristics and provided that the background level of hybridization remains low. Usually the solid substrate will be a microtiter plate, a membrane (e.g. nylon or nitrocellulose) or a microsphere (bead). Prior to application to the membrane or fixation it may be convenient to modify the nucleic acid probe in order to facilitate fixation or improve the hybridization efficiency. Such modifications may encompass homopolymer tailing, coupling with different reactive groups such as aliphatic

groups, NH_2 groups, SH groups, carboxylic groups, or coupling with biotin or haptens.

The present invention also relates to a diagnostic kit for use in determining the genotype of HCV, said kit comprising a primer as defined above.

The present invention also relates to a diagnostic kit for use in determining the genotype of HCV, said kit comprising a probe as defined above.

The present invention also relates to a diagnostic kit as defined above, wherein said probe(s) is(are) attached to a solid substrate.

The present invention also relates to a diagnostic kit as defined above, wherein a range of said probes is attached to specific locations on a solid substrate.

The present invention also relates to a diagnostic kit as defined above, wherein said solid support is a membrane strip and said probes are coupled to the membrane in the form of parallel lines.

The present invention also relates to a method for the detection of HCV nucleic acids present in a biological sample, comprising:

(i) possibly extracting sample nucleic acid,

(ii) amplifying the nucleic acid with at least one primer as defined above,

(iii) detecting the amplified nucleic acids.

The present invention also relates to a method for the detection of HCV nucleic acids present in a biological sample, comprising:

(i) possibly extracting sample nucleic acid,

(ii) possibly amplifying the nucleic acid with at least one primer as defiend above, or with a universal HCV primer,

(iii) hybridizing the nucleic acids of the biological sample, possibly under denatured conditions, at appropriate conditions with one or more probes as defined above, with said probes being preferably attached to a solid substrate,

(iv) possibly washing at appropriate conditions,

(v) detecting the hybrids formed.

The present invention also relates to a method for detecting the presence of one or more HCV genotypes present in a biological sample, comprising:

(i) possibly extracting sample nucleic acid,

(ii) specifically amplifying the nucleic acid with at least one primer as defined

above,

(iii) detecting said amplified nucleic acids.

The present invention also relates to a method for detecting the presence of one or more HCV genotypes present in a biological sample, comprising:

(i) possibly extracting sample nucleic acid,

(ii) possibly amplifying the nucleic acid with at least one primer as defined above or with a universal HCV primer,

(iii) hybridizing the nucleic acids of the biological sample, possibly under denatured conditions, at appropriate conditions with one or more probes as defined above, with said probes being preferably attached to a solid substrate,

(iv) possibly washing at appropriate conditions,

(v) detecting the hybrids formed,

(vi) inferring the presence of one or more HCV genotypes present from the observed hybridization pattern.

The present invention also relates to a method as defined above, wherein said probes are further characterized as defined above.

The present invention also relates to a method as defined above, wherein said nucleic acids are labelled during or after amplification.

Preferably, this technique could be performed in the 5' non-coding, Core or NS5B region.

The term "nucleic acid" can also be referred to as analyte strand and corresponds to a single- or double-stranded nucleic acid molecule. This analyte strand is preferentially positive- or negative stranded RNA, cDNA or amplified cDNA.

The term "biological sample" refers to any biological sample (tissue or fluid) containing HCV nucleic acid sequences and refers more particularly to blood serum or plasma samples.

The term "universal HCV primer" refers to oligonucleotide sequences complementary to any of the conserved regions of the HCV genome.

The expression "appropriate" hybridization and washing conditions are to be understood as stringent and are generally known in the art (e.g. Maniatis et al., Molecular Cloning: A Laboratory Manual, New York, Cold Spring Harbor Laboratory, 1982).

However, according to the hybridization solution (SSC, SSPE, etc.), these probes should be hybridized at their appropriate temperature in order to attain sufficient specificity.

The term "labelled" refers to the use of labelled nucleic acids. This may include the use of labelled nucleotides incorporated during the polymerase step of the amplification such as illustrated by Saiki et al. (1988) or Bej et al. (1990) or labelled primers, or by any other method known to the person skilled in the art.

The process of the invention comprises the steps of contacting any of the probes as defined above, with one of the following elements:

- either a biological sample in which the nucleic acids are made available for hybridization,
- or the purified nucleic acids contained in the biological sample
- or a single copy derived from the purified nucleic acids,
- or an amplified copy derived from the purified nucleic acids, with said elements or with said probes being attached to a solid substrate.

The expression "inferring the presence of one or more HCV genotypes present from the observed hybridization pattern" refers to the identification of the presence of HCV genomes in the sample by analyzing the pattern of binding of a panel of oligonucleotide probes. Single probes may provide useful information concerning the presence or absence of HCV genomes in a sample. On the other hand, the variation of the HCV genomes is dispersed in nature, so rarely is any one probe able to identify uniquely a specific HCV genome. Rather, the identity of an HCV genotype may be inferred from the pattern of binding of a panel of oligonucleotide probes, which are specific for (different) segments of the different HCV genomes. Depending on the choice of these oligonucleotide probes, each known HCV genotype will correspond to a specific hybridization pattern upon use of a specific combination of probes. Each HCV genotype will also be able to be discriminated from any other HCV genotype amplified with the same primers depending on the choice of the oligonucleotide probes. Comparison of the generated pattern of positively hybridizing probes for a sample containing one or more unkown HCV sequences to a scheme of expected hybridization patterns, allows one to clearly infer the HCV genotypes present in said sample.

The present invention thus relates to a method as defined above, wherein one

or more hybridization probes are selected from any of SEQ ID NO 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103 or 105 or sequence variants thereof as defined above.

In order to distinguish the amplified HCV genomes from each other, the target polynucleic acids are hybridized to a set of sequence-specific DNA probes targetting HCV genotypic regions (unique regions) located in the HCV polynucleic acids.

Most of these probes target the most type- or subtype-specific regions of HCV genotypes, but some can be caused to hybridize to more than one HCV genotype.

According to the hybridization solution (SSC, SSPE, etc.), these probes should be stringently hybridized at their appropriate temperature in order to attain sufficient specificity. However, by slightly modifying the DNA probes, either by adding or deleting one or a few nucleotides at their extremities (either 3' or 5'), or substituting some non-essential nucleotides (i.e. nucleotides not essential to discriminate between types) by others (including modified nucleotides or inosine) these probes or variants thereof can be caused to hybridize specifically at the same hybridization conditions (i.e. the same temperature and the same hybridization solution). Also changing the amount (concentration) of probe used may be beneficial to obtain more specific hybridization results. It should be noted in this context, that probes of the same length, regardless of their GC content, will hybridize specifically at approximately the same temperature in TMACl solutions (Jacobs et al., 1988).

Suitable assay methods for purposes of the present invention to detect hybrids formed between the oligonucleotide probes and the nucleic acid sequences in a sample may comprise any of the assay formats known in the art, such as the conventional dot-blot format, sandwich hybridization or reverse hybridization. For example, the detection can be accomplished using a dot blot format, the unlabelled amplified sample being bound to a membrane, the membrane being incorporated with at least one labelled probe under suitable hybridization and wash conditions, and the presence of bound probe being monitored.

An alternative and preferred method is a "reverse" dot-blot format, in which the amplified sequence contains a label. In this format, the unlabelled oligonucleotide probes are bound to a solid support and exposed to the labelled sample under appropriate stringent hybridization and subsequent washing conditions. It is to be

understood that also any other assay method which relies on the formation of a hybrid between the nucleic acids of the sample and the oligonucleotide probes according to the present invention may be used.

According to an advantageous embodiment, the process of detecting one or more HCV genotypes contained in a biological sample comprises the steps of contacting amplified HCV nucleic acid copies derived from the biological sample, with oligonucleotide probes which have been immobilized as parallel lines on a solid support.

According to this advantageous method, the probes are immobilized in a Line Probe Assay (LiPA) format. This is a reverse hybridization format (Saiki et al., 1989) using membrane strips onto which several oligonucleotide probes (including negative or positive control oligonucleotides) can be conveniently applied as parallel lines.

The invention thus also relates to a solid support, preferably a membrane strip, carrying on its surface, one or more probes as defined above, coupled to the support in the form of parallel lines.

The LiPA is a very rapid and user-friendly hybridization test. Results can be read after 4 hours. after the start of the amplification. After amplification during which usually a non-isotopic label is incorporated in the amplified product, and alkaline denaturation, the amplified product is contacted with the probes on the membrane and the hybridization is carried out for about 1 to 1,5 h hybridized polynucleic acid is detected. From the hybridization pattern generated, the HCV type can be deduced either visually, but preferably using dedicated software. The LiPA format is completely compatible with commercially available scanning devices, thus rendering automatic interpretation of the results very reliable. All those advantages make the LiPA format liable for the use of HCV detection in a routine setting. The LiPA format should be particularly advantageous for detecting the presence of different HCV genotypes.

The present invention also relates to a method for detecting and identifying novel HCV genotypes, different from the known HCV genomes, comprising the steps of:

- determining to which HCV genotype the nucleotides present in a biological sample belong, according to the process as defined above,
- in the case of observing a sample which does not generate a hybridization

pattern compatible with those defined in Table 3, sequencing the portion of the HCV genome sequence corresponding to the aberrantly hybridizing probe of the new HCV genotype to be determined.

The present invention also relates to a method for preparing a polynucleic acid according to the present invention. These methods include any method known in the art for preparing polynucleic acids (e.g. the phosphodiester method for synthesizing oligonucleotides as described by Agarwal et al. 1972, Agnew. Chem. Int. Ed. Engl. 11:451, the phosphotriester method of Hsiung et al. 1979, Nucleic Acid Res. 6:1371, or the automated diethylphosphoramidite method of Baeucage et al. 1981, Tetrahedron Letters 22:1859-1862.). Alternatively, the polynucleic acids of the present invention may be isolated fragments of naturally occuring or cloned DNA or RNA. In addition, the oligonucleotides according to the present invention may be synthesized automatically on commercial instruments sold by a variety of manufacturers.

The present invention particularly also relates to a polypeptide having an amino acid sequence encoded by a polynucleic acid as defined above, or a part thereof which is unique to at least one of the HCV subtypes or types as defined in Table 5, and which contains at least one amino acid differing from any of the known HCV types or subtypes, or an analog thereof being substantially homologous and biologically equivalent .

The term 'polypeptide' refers to a polymer of amino acids and does not refer to a specific length of the product; thus, peptides, oligopeptides, and proteins are included within the definition of polypeptide. This term also does not refer to or exclude post-expression modifications of the polypeptide, for example, glycosylations, acetylations, phosphorylations and the like. Included within the definition are, for example, polypeptides containing one or more analogues of an amino acid (including, for example, unnatural amino acids, PNA, etc.), polypeptides with substituted linkages, as well as other modifications known in the art, both naturally occurring and non-naturally occurring.

The term "unique" is referred above.

By "biologically equivalent" as used throughout the specification and claims, it is meant that the compositions are immunogenically equivalent to the proteins (polypeptides) or peptides of the invention as defined above and below.

By "substantially homologous" as used throughout the ensuing specification and claims to describe proteins and peptides, it is meant a degree of homology in the amino acid sequence to the proteins or peptides of the invention. Preferably the degree of homology is in excess of 90, preferably in excess of 95, with a particularly preferred group of proteins being in excess of 99 homologous with the proteins or peptides of the invention.

The term "analog" as used throughout the specification or claims to describe the proteins or peptides of the present invention, includes any protein or peptide having an amino acid residue sequence substantially identical to a sequence specifically shown herein in which one or more residues have been conservatively substituted with a biologically equivalent residue. Examples of conservative substitutions include the substitution of one-polar (hydrophobic) residue such as isoleucine, valine, leucine or methionine for another, the substitution of one polar (hydrophillic) residue for another such as between arginine and lysine, between glutamine and asparagine, between glycine and serine, the substitution of one basic residue such as lysine, arginine or histidine for another, or the substitution of one acidic residue, such as aspartic acid or glutamic acid for another. Examples of allowable mutations acccording to the present inevntion can be found in Table 4.

The phrase "conservative substitution" also includes the use of a chemically derivatized residue in place of a non-derivatized residue provided that the resulting protein or peptide is biologically equivalent to theprotein or peptide of the invention.

"Chemical derivative" refers to a protein or peptide having one or more residues chemically derivatized by reaction of a functional side group. Examples of such derivatized molecules, include but are not limited to, those molecules in which free amino groups have been derivatized to form amine hydrochlorides, p-toluene sulfonyl groups, carbobenzoxy groups, t-butyloxycarbonyl groups, chloracetyl groups or formyl groups. Free carboxyl groups may be derivatized to form salts, methyl and ethyl esters or other types of esters or hydrazides. Free hydroxyl groups may be derivatized to form O-acyl or O-alkyl derivatives. The imidazole nitrogen of histidine may be derivatized to form N-imbenzylhistidine. Also included as chemical derivatives are those proteins or peptides which contain one or more naturally-occurring amino acid derivatives of the twenty standard amino acids. For examples : 4-hydroxyproline may be substituted for proline; 5-hydroxylysine may be substituted for lysine; 3-

methylhistidine may be substituted for histidine; homoserine may be substituted for serine; and ornithine may be substituted for lysine. The proteins or peptides of the present invention also include any protein or peptide having one or more additions and/or deletions or residues relative to the sequence of a peptide whose sequence is shown herein, so long as the peptide is biologically equivalent to the proteins or peptides of the invention.

It is to be noted that, at the level of the amino acid sequence, at least one amino acids difference (with respect to known HCV amino acid sequences) is sufficient to be part of the invention, which means that the polypeptides of the invention correspond to polynucleic acids having at least one nucleotide difference (with known HCV polynucleic acid sequences) involving an amino acid difference in the encoded polyprotein.

As the NS4 and the Core regions are known to contain several epitopes, for example characterized in patent application EP-A-0 489 968, and as the E1 protein is expected to be subject to immune attack as part of the viral envelope and expected to contain epitopes, the NS4, Core and E1 epitopes of the new types and subtypes disclosed herein will consistently differ from the epitopes present in previously known genotypes. This is examplified by the type-specificity of NS4 synthetic peptides as described in Simmonds et al. (1993c) and Stuyver et al. (1993b) and PCT/EP 94/01323 and the type-specificity of recombinant E1 proteins as described in Maertens et al. (1994).

The peptides according to the present invention contain preferably at least 3, preferably 4, 5 contiguous HCV amino acids, 6, 7 preferably however at least 8 contiguous HCV amino acids, at least 10 or at least 15 (for instance at least 9, 10, 11, 12, 13, 14, 15, 16, 17, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50 or more amino acids).

TABLE 4

Amino acids	Synonymous groups
Ser (S)	Ser, Thr, Gly, Asn
Arg (R)	Arg, His, Lys, Glu, Gln

Leu (L)	Leu; Ile, Met, Phe, Val, Tyr
Pro (P)	Pro, Ala, Thr, Gly
Thr (T)	Thr, Pro, Ser, Ala, Gly, His, Gln
Ala (A)	Ala, Pro, Gly, Thr
Val (V)	Val, Met, Ile, Tyr, Phe, Leu, Val
Gly (G)	Gly, Ala, Thr, Pro, Ser
Ile (I)	Ile, Met, Leu, Phe, Val, Ile, Tyr
Phe (F)	Phe, Met, Tyr, Ile, Leu, Trp, Val
Tyr (Y)	Tyr, Phe, Trp, Met, Ile, Val, Leu
Cys (C)	Cys, Ser, Thr, Met
His (H)	His, Gln, Arg, Lys, Glu, Thr
Gln (Q)	Gln, Glu, His, Lys, Asn, Thr, Arg
Asn (N)	Asn, Asp, Ser, Gln
Lys (K)	Lys, Arg, Glu, Gln, His
Asp (D)	Asp, Asn, Glu, Gln
Glu (E)	Glu, Gln, Asp, Lys, Asn, His, Arg
Met (M)	Met, Ile, Leu, Phe, Val

Table 4 Overview of the amino acid substitutions which could form the basis of analogs (muteins) as defined above

The polypeptides of the invention, and particularly the fragments, can be prepared by classical chemical synthesis.

The synthesis can be carried out in homogeneous solution or in solid phase.

For instance, the synthesis technique in homogeneous solution which can be used is the one described by Houbenweyl in the book entitled "Methode der organischen chemie" (Method of organic chemistry) edited by E. Wunsh, vol. 15-I et II. THIEME, Stuttgart 1974.

The polypeptides of the invention can also be prepared in solid phase according to the methods described by Atherton and Shepard in their book entitled "Solid phase peptide synthesis" (IRL Press, Oxford, 1989).

The polypeptides according to this invention can be prepared by means of recombinant DNA techniques as described by Maniatis et al., Molecular Cloning: A Laboratory Manual, New York, Cold Spring Harbor Laboratory, 1982).

The present invention relates particularly to a polypeptide as defined above, comprising in its amino acid sequence at least one of the following amino acid residues:

I15, C38, V44, A49, Q43, P49, Q55, A58, S60 or D60, E68 or V68, H70, A71 or Q71 or N71, D72, H81, H101, D106, S110, L130, I134, E135, L140, S148, T150 or E150, Q153, F155, D157, G160, E165, I169, F181, L186, T190, T192 or I192 or H192, I193, A195, S196, R197 or N197 or K197, Q199 or D199 or H199, N199, F200 or T200, A208, I213, M216 or S216, N217 or S217 or G217 or K217, T218, I219, A222, Y223, I230, W231 or L231, S232 or H232 or A232, Q233, E235 or L235, F236 or T236, F237, L240 or M240, A242, N244, N249, I250 or K250 or R250, A252 or C252, A254, I255 or V255, D256 or M256, E257, E260 or K260, R261, V268, S272 or R272, I285, G290 or F290, A291, A293 or L293 or W293, T294 or A294, S295, H295, K296 or E296, Y297 or M297, I299 or Y299, I300, S301, P316, S2646, A2648, G2649, A2650, V2652, Q2653, H2656 or L2656, D2657, F2659, K2663 or Q2663, A2667 or V2667, D2677, L2681, M2686 or Q2686 or E2686, A2692 or K2692, H2697, I2707, L2708 or Y2708, A2709, A2719 or M2719, F2727, T2728 or D2728, E2729, F2730 or Y2730, I2741, I2745, V2746 or E2746 or L2746 or K2746, A2748, S2749 or P2749, R2750, E2751, D2752 or N2752 or S2752 or T2752 or V2752 or I2752 or Q2752, S2753 or D2753 or G2753, D2754, A2755, L2756 or Q2756, or R2757,

with said notation being composed of a letter representing the amino acid residue by its one-letter code, and a number representing the amino acid numbering according to Kato et al., 1990 as shown in Table 1 (see also the numbering in Figures 2, 4 and 6),

or a part thereof which is unique to at least one of the HCV subtypes or types as defined in Table 5, and which contains at least one amino acid differing from any of the known HCV types or subtypes, or an analog thereof being substantially homologous and biologically equivalent to said polypeptide or part thereof.

These unique amino acid residues can be deduced from aligning the new HCV amino acid sequences as given in Figure 3 to all known HCV sequences. An alignment with the new sequences as represented in SEQ ID NO 1 to 106 is given in for instance Figures 2, 4 and 6. It should be clear that the alignments given in these figures may be completed with all known HCV sequences to illustrate that any of the above-given unique residues is indeed unique for at least one of the new HCV sequences of the present invention.

Within the group of unique and new amino acid residues of the present

invention, unique residues may be found which are specific for the following new types (subtypes) of HCV according to the HCV classification system used in the present invention: type 1 subtype 1d, 1e, 1f or 1g isolates; type 2 subtype 2e, 2f, 2g, 2h, 2i, 2k or 2l isolates; type 3 subtype 3g isolates; type 4 subtype 4k, 4l or 4m isolates; type 7 subtype 7a, 7c or 7d isolates, type 9, type 10 or type 11 isolates. In order to obtain these residues the alignments given in Figures 2, 4 and 6 may be used to deduce the type- and or subtype-specificity of any of the unique residues given above.

For example T190 (detected in subtype 1d) refers to a threomine at position 190 (see Figure 2). In other sequences only a serine (S190) or exceptionally an alanine (A190 in type 10a) can be detected.

The polypeptides according to this embodiment of the invention may be possibly labelled, or attached to a solid substrate, or coupled to a carrier molecule such as biotin, or mixed with a proper adjuvant all known in the art and according to the intended use (diagnostic, therapeutic or prophylactic).

The present invention also relates to a polypeptide as defined above, comprising in its amino acid sequence at least one of the sequences repesented by SEQ ID NO107 to 207 as listed above, or a part thereof which is unique to at least one of the HCV subtypes or types as defined in Table 5, or an analog thereof being substantially homologous and biologically equivalent to said polypeptide or part thereof.

The present invention relates also to a polypeptide having an amino acid sequence as represented in any of SEQ ID NO 1 to 106, or a part thereof which is unique to at least one of the HCV subtypes or types as defined in Table 5, or an analog thereof being substantially homologous and biologically equivalent to said polypeptide or part thereof.

The variable region in the core protein (V-CORE in Fig. 2) has been shown to be useful for serotyping (Machida et al., 1992). The sequence of the type 1 subtype 1d, 1e, 1f or 1g sequence; type 2 subtype 2e, 2f, 2g, 2h, 2i, 2k and 2l sequence; type 3 subtype 3g; type 4, subtype 4k, 4l or 4m sequence; type 7 (subtype 7a, 7c and 7d sequences), 9, 10 or 11 sequences of the present invention show type-specific features in this region. The peptide from amino acid 68 to 78 (V-core region) shows the following unique sequence for the sequences of the present invention (see

figure 2):

ARQSDGRSWAQ or ARRSEGRSWAQ as for subtype 1d (SEQ ID NO 107 and 108)

ERRPEGRSWAQ as for subtype 1e (SEQ ID NO 109)

ARRPEGRSWAQ as for subtype 1f (SEQ ID NO 110)

DRRTTGKSWGR as for subtype 2k (SEQ ID NO 111)

DRRATGRSWGR as for subtype 2e (SEQ ID NO 112)

DRRATGKSWGR as for subtype 2f (SEQ ID NO 113)

VRQPTGRSWGQ as for type 9 (SEQ ID NO 114)

VRHQTGRTWAQ as for subtype 7a and 7c (SEQ ID NO 115)

VRQNQGRTWAQ as for subtype 7d (SEQ ID NO 116)

ARRTEGRSWAQ as for type 10 (SEQ ID NO 117)

VRRTTGRXXXX or VRRTTGRTWAQ as for type 11 (SEQ ID NO 118 and 119)

Five type-specific variable regions (V1 to V5) can be identified after aligning E1 amino acid sequences of the genotypes of the present invention to the genotypes already known, as shown in Figure 2.

Region V1 encompasses amino acids 192 to 203, this is the amino-terminal 10 amino acids of the E1 protein. The following unique sequences as shown in Fig. 2 can be deduced:

HEVRNASGVYHV or HEVRNASGVYHL as for subtype 1d, (SEQ ID NO 120 and 121)

YEVHSTTDGYHV as for subtype 1f (SEQ ID NO 122)

VEVKNTSQAYMA as for subtype 2e (SEQ ID NO 123)

IQVKNNSHFYMA as for subtype 2f (SEQ ID NO 124)

VQVKNTSTMYMA as for subtype 2g (SEQ ID NO 125)

VQVKNTSHSYMV as for subtype 2h (SEQ ID NO 126)

VQVANRSGSYMV as for subtype 2i (SEQ ID NO 127)

VEIKNTXNTYVL or VEIKNTSNTYVL as for subtype 2k (SEQ ID NO 128 and 129)

INYRNVSGIYYV or INYRNTSGIYHV or INYHNTSGIYHI or TNYRNVSGIYHV a for subtype 4k (SEQ ID NO 130, 131, 132 or 133)

QHYRNVSGIYHV as for subtype 4l (SEQ ID NO 134)

IQVKNASGIYHL as for type 9 (SEQ ID NO 135)

AHYTNKSGLYHL as for subtype 7c (SEQ ID NO 136)

LNYANKSGLYHL as for subtype 7d (SEQ ID NO 137)

LEYRNASGLYMV as for type 10 (SEQ ID NO 138)

Region V2 encompasses amino acids 213 to 223. The following unique sequences can be found in the V2 region as shown in Figure 2:

IYEMDGMIMHY or IYEMSGMILHA as for subtype 1d, (SEQ ID NO 139 and 140)

VYEAKDIILHT as for subtype 1f (SEQ ID NO 141)

VWQLXDAVLHV as for subtype 2e (SEQ ID NO 142)

VWQLRDAVLHV as for subtype 2f (SEQ ID NO 143)

IWQMQGAVLHV as for subtupe 2g (SEQ ID NO 144)

VWQLKDAVLHV as for subtype 2h (SEQ ID NO 145)

VWQLEEAVLHV as for subtype 2i (SEQ ID NO 146)

TWQLXXAVLHV as for subtype 2k (SEQ ID NO 147)

VYEADHHILHL or VYEADHHILAL or VFEADHHILHL as for subtype 4k (SEQ ID NO 148, 149 and 150)

VYESDHHILHL as for subtype 4l (SEQ ID NO 151)

VFEAETMILHL as for type 9 (SEQ ID NO 152)

VYEAETLILHL as for subtype 7c (SEQ ID NO 153)

VYEANGMILHL as for subtype 7d (SEQ ID NO 154)

VYEAGDIILHL as for type 10. (SEQ ID NO 155)

Region V3 encompasses the amino acids 230 to 242. The following unique V3 region sequences can be deduced from Figure 2:

VREDNHLRCWMAL or VRENNSSRCWMAL as for subtype 1d (SEQ ID NO 156 and 157)

IREGNISRCWVLP as for subtype 1f (SEQ ID NO 158)

ENSSGRFHCWIPI as for subtype 2e (SEQ ID NO 159)

ERSGNRTFCWTAV as for subtype 2f (SEQ ID NO 160)

ELQGNKSRCWIPV as for subtype 2g (SEQ ID NO 162)

ERHQNQSRCWIPV as for subtype 2h (SEQ ID NO 163)

EWKDNTSRCWIPV as for subtype 2i (SEQ ID NO 164)

EREGNSSRCWIPV as for subtype 2k (SEQ ID NO 165)

VREGNQSRCWVAL or VRTGNQSRCWVAL or VRVGNQSSCWVAL or VRVGNQSRCWVAL or VKEGNHSRCWVAL as for subtype 4k (SEQ ID NO 166, 167, 168 or 169)

VKTGNTSRCWVAL as for subtype 4l (SEQ ID NO 170)

IKAGNESRCWLPV as for type 9 (SEQ ID NO 171)

VKXXNQSRCWVQA as for subtype 7c (SEQ ID NO 172)

VKTGNLTKCWLSA as for subtype 7d (SEQ ID NO 173)

VRSGNTSRCWIPV as for type 10 (SEQ ID NO 174)

Region V4 encompasses the amino acids 248 to 257. The following unique V4 region sequences can be deduced from figure 2:

VKNASVPTAA or VKDANVPTAA as for subtype 1d (SEQ ID NO 175 and 176)

ARIANAPIDE as for subtype 1f (SEQ ID NO 177)

VSKPGALTKG as for subtype 2e (SEQ ID NO 178)

VSRPGALTRG as for subtype 2f (SEQ ID NO 179)

VNQPGALTRG as for subtype 2g (SEQ ID NO 180)

VSQPGALTRG as for subtype 2h (SEQ ID NO 181)

VSQPGALTKG as for subtype 2i (SEQ ID NO 182)

VSRPGALTEG as for subtype 2k (SEQ ID NO 183)

APYIGAPLES or APYTAAPLES as for subtype 4k (SEQ ID NO 184 and 185)

APILSAPLMS as for subtype 4l (SEQ ID NO 186)

VPNSSVPIHG as for type 9 (SEQ ID NO 187)

VPNASTPVTG as for subtype 7c (SEQ ID NO 188)

VQNASVSIRG as for subtype 7d (SEQ ID NO 189)

VKSPCAATAS as for type 10 (SEQ ID NO 190)

Region V5 encompasses the amino acids 294 to 303. The following unique V5 region peptides can be deduced from figure 2:

SPRMHHTTQE or SPRLYHTTQE as for subtype 1d (SEQ ID NO 191 and 192)

TSRRHWTVQD as for subtype 1f (SEQ ID NO 193)

APKRHYFVQE as for subtype 2e (SEQ ID NO 194)

SPQYHTFVQE as for subtype 2f (SEQ ID NO 195)

SPQHHNFSQD as for subtype 2g (SEQ ID NO 196)

SPQHHIFVQD as for subtype 2h (SEQ ID NO 197)

SPEHHHFVQD as for subtype 2k (SEQ ID NO 198)

RPRRHWTTQD or RPRRHWTAQD or QPRRHWTTQD or RPRRHWTTQE as for subtype 4k (SEQ ID NO 199, 200, 201 or 202)

QPRRHWTVQD as for subtype 4l (SEQ ID NO 203)

RPKYHQVTQD as for type 9 (SEQ ID NO 204)

RPRMHQVVQE as for subtype 7c (SEQ ID NO 205)

RPRMYEIAQD as for subtype 7d (SEQ ID NO 206)

RHRQHWTVQD as for type 10 (SEQ ID NO 207)

The above given list of peptides are particularly useful for treatment and vaccine and diagnostic development.

Also comprised in the present invention is any synthetic peptide (see below) or polypeptide containing at least an epitope derived from the above-defined peptides in their peptidic chain. Also comprised within the present invention is any synthetic peptide or polypeptide comprising at least 6, 7, 8, or 9 contiguous amino acids derived from the above-defined peptides in their peptidic chain.

As used herein, 'epitope' or 'antigenic determinant' means an amino acid sequence that is immunoreactive. Generally an epitope consists of at least 3 to 4 amino acids, and more usually, consists of at least 5 or 6 amino acids, sometimes the epitope consists of about 7 to 8, or even about 10 amino acids.

The present invention particularly relates to any peptide (see below) or polypeptide contained in any of the amino acid sequences as represented in SEQ ID NO 2, 4, 7, 9, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104 or 106 (see Table 5 and Figure 3, Examples section).

The present invention also relates to a recombinant polypeptide encoded by a polynucleic acid as defined above, or a part thereof which is unique to any of the HCV subtypes or types as defined in Table 5, or an analog thereof being substantially homologous and biologically equivalent to said polypeptide.

The present invention also relates to a recombinant expression vector comprising a polynucleic acid or a part thereof as defined above, operably linked to prokaryotic, eukaryotic or viral transcription and translation control elements.

In general said recombinant vector will comprise a vector sequence, an appropriate prokaryotic, eukaryotic or viral promoter sequence followed by the nucleotide sequences as defined above, with said recombinant vector allowing the expression of any one of the polypeptides as defined above in a prokaryotic, or eukaryotic host or in living mammals when injected as naked DNA, and more particularly a recombinant vector allowing the expression of any of the new HCV sequences of the invention spanning particularly the following amino acid positions:

- a polypeptide starting in the region between positions 1 and 10 and ending at any position in the region between positions 70 and 420, more particularly a polypeptide spanning positions 1 to 70, 1 to 85, positions 1 to 120, positions 1 to 150, positions 1 to 191, or positions 1 to 200, for expression of the Core protein, and a polypeptide spanning positions 1 to 263, positions 1 to 326, positions 1 to 383, or positions 1 to 420 for expression of the Core and E1 protein;
- a polypeptide starting at any position in the region between positions 117 and 192, and ending at any position in the region between positions 263 and 420, for expression of E1, or forms that have the hydrophobic region deleted (positions 264 to 293 plus or minus 8 amino acids);
- a polypeptide starting at any position in the region between positions 1556 and 1688, and ending at any position in the region between positions 1739 and 1764, for expression of NS4, more particularly ;a polypeptide starting at position 1658 and ending at position 1711, for expression of NS4a antigen, and more particularly, a polypeptide starting at position 1712 and ending in the region between positions 1743 and 1972 (for instance 1712-1743, 1712-1764, 1712-1782, 1712-1972, 1712-1782, 1712-1902), for expression of NS4b antigen or parts thereof.

Any other HCV vector construction known in the art may also be used for the recombinant polypeptides of the present invention.

Also any of the known purification methods for recombinant proteins may be used for the production of the recombinant polypeptides of the present invention, particularly the HCV recombinant polypeptide purification methods as disclosed in PCT/EP 95/03031 in name of Innogenetics N.V.

The term "vector" may comprise a plasmid, a cosmid, a phage, or a virus or

a transgenic animal. Particularly useful for vaccine development may be BCG or adenoviral vectors, as well as avipox recombinant viruses.

The present invention also relates to a method for the production of a recombinant polypeptide as defined above, comprising:

- transformation of an appropriate cellular host with a recombinant vector, in which a polynucleic acid or a part thereof according to as defined above has been inserted under the control of appropriate regulatory elements,
- culturing said transformed cellular host under conditions enabling the expression of said insert, and,
- harvesting said polypeptide.

The term 'recombinantly expressed' used within the context of the present invention refers to the fact that the proteins of the present invention are produced by recombinant expression methods be it in prokaryotes, or lower or higher eukaryotes as discussed in detail below.

The term 'lower eukaryote' refers to host cells such as yeast, fungi and the like. Lower eukaryotes are generally (but not necessarily) unicellular. Preferred lower eukaryotes are yeasts, particularly species within Saccharomyces, Schizosaccharomyces, Kluveromyces, Pichia (e.g. Pichia pastoris), Hansenula (e.g. Hansenula polymorpha), Yarowia, Schwaniomyces, Schizosaccharomyces, Zygosaccharomyces and the like. Saccharomyces cerevisiae, S. carlsbergensis and K. lactis are the most commonly used yeast hosts, and are convenient fungal hosts.

The term 'prokaryotes' refers to hosts such as E.coli, Lactobacillus, Lactococcus, Salmonella, Streptococcus, Bacillus subtilis or Streptomyces. Also these hosts are contemplated within the present invention.

The term 'higher eukaryote' refers to host cells derived from higher animals, such as mammals, reptiles, insects, and the like. Presently preferred higher eukaryote host cells are derived from Chinese hamster (e.g. CHO), monkey (e.g. COS and Vero cells), baby hamster kidney (BHK), pig kidney (PK15), rabbit kidney 13 cells (RK13), the human osteosarcoma cell line 143 B, the human cell line HeLa and human hepatoma cell lines like Hep G2, and insect cell lines (e.g. Spodoptera frugiperda). The host cells may be provided in suspension or flask cultures, tissue cultures, organ cultures and the like. Alternatively the host cells may also be transgenic animals.

The term 'recombinant polynucleotide or nucleic acid' intends a polynucleotide

or nucleic acid of genomic, cDNA, semisynthetic, or synthetic origin which, by virtue of its origin or manipulation : (1) is not associated with all or a portion of a polynucleotide with which it is associated in nature, (2) is linked to a polynucleotide other than that to which it is linked in nature, or (3) does not occur in nature.

The term 'recombinant host cells', 'host cells', 'cells', 'cell lines', 'cell cultures', and other such terms denoting microorganisms or higher eukaryotic cell lines cultured as unicellular entities refer to cells which can be or have been, used as recipients for a recombinant vector or other transfer polynucleotide, and include the progeny of the original cell which has been transfected. It is understood that the progeny of a single parental cell may not necessarily be completely identical in morphology or in genomic or total DNA complement as the original parent, due to natural, accidental, or deliberate mutation.

The term 'replicon' is any genetic element, e.g., a plasmid, a chromosome, a virus, a cosmid, etc., that behaves as an autonomous unit of polynucleotide replication within a cell; i.e., capable of replication under its own control.

The term 'vector' is a replicon further comprising sequences providing replication and/or expression of a desired open reading frame.

The term 'control sequence' refers to polynucleotide sequences which are necessary to effect the expression of coding sequences to which they are ligated. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include promoter, ribosomal binding site, splicing sites and terminators; in eukaryotes, generally, such control sequences include promoters, splicing sites, terminators and, in some instances, enhancers. The term 'control sequences' is intended to include, at a minimum, all components whose presence is necessary for expression, and may also include additional components whose presence is advantageous, for example, leader sequences which govern secretion.

The term 'promoter' is a nucleotide sequence which is comprised of consensus sequences which allow the binding of RNA polymerase to the DNA template in a manner such that mRNA production initiates at the normal transcription initiation site for the adjacent structural gene.

The expression 'operably linked' refers to a juxtaposition wherein the components so described are in a relationship permitting them to function in their

intended manner. A control sequence 'operably linked' to a coding sequence is ligated in such a way that expression of the coding sequence is achieved under conditions compatible with the control sequences.

The segment of the HCV cDNA encoding the desired sequence inserted into the vector sequence may be attached to a signal sequence. Said signal sequence may be that from a non-HCV source, e.g. the IgG or tissue plasminogen activator (tpa) leader sequence for expression in mammalian cells, or the α-mating factor sequence for expression into yeast cells, but particularly preferred constructs according to the present invention contain signal sequences appearing in the HCV genome before the respective start points of the proteins.

A variety of vectors may be used to obtain recombinant expression of HCV single or specific oligomeric envelope proteins of the present invention. Lower eukaryotes such as yeasts and glycosylation mutant strains are typically transformed with plasmids, or are transformed with a recombinant virus. The vectors may replicate within the host independently, or may integrate into the host cell genome.

Higher eukaryotes may be transformed with vectors, or may be infected with a recombinant virus, for example a recombinant vaccinia virus. Techniques and vectors for the insertion of foreign DNA into vaccinia virus are well known in the art, and utilize, for example homologous recombination. A wide variety of viral promoter sequences, possibly terminator sequences and poly(A)-addition sequences, possibly enhancer sequences and possibly amplification sequences, all required for the mammalian expression, are available in the art. Vaccinia is particularly preferred since vaccinia halts the expression of host cell proteins. Vaccinia is also very much preferred since it allows the expression of f.i. E1 and E2 proteins of HCV in cells or individuals which are immunized with the live recombinant vaccinia virus. For vaccination of humans the avipox and Ankara Modified Virus (AMV) are particularly useful vectors.

Also known are insect expression transfer vectors derived from baculovirus Autographa californica nuclear polyhedrosis virus (AcNPV), which is a helper-independent viral expression vector. Expression vectors derived from this system usually use the strong viral polyhedrin gene promoter to drive the expression of heterologous genes. Different vectors as well as methods for the introduction of heterologous DNA into the desired site of baculovirus are available to the man skilled

in the art for baculovirus expression. Also different signals for posttranslational modification recognized by insect cells are known in the art.

The present invention also relates to a host cell transformed with a recombinant vector as defined above.

The present invention also relates to a method for detecting antibodies to HCV present in a biological sample, comprising:

(i) contacting the biological sample to be analysed for the presence of HCV with a polypeptide as defined above,

(ii) detecting the immunological complex formed between said antibodies and said polypeptide.

The present invention also relates to a method for HCV typing, comprising:

(i) contacting the biological sample to be analysed for the presence of HCV with a polypeptide as defined above,

(ii) detecting the immunological complex formed between said antibodies and said polypeptide.

The present invention also relates to a diagnostic kit for use in detecting the presence of HCV, said kit comprising at least one polypeptide as defined above, with said polypeptide being preferably bound to a solid support.

The present invention also relates to a diagnostic kit for HCV typing, said kit comprising at least one polypeptide as defined above, with said polypeptide being preferably bound to a solid support.

The present invention also relates to diagnostic kit according as defined above, said kit comprising a range of said polypeptides which are attached to specific locations on a solid substrate.

The present invention also relates to a diagnostic kit as defined above, wherein said solid support is a membrane strip and said polypeptides are coupled to the membrane in the form of parallel lines.

The immunoassay methods according to the present invention may utilize antigens from the different domains of the new and unique polypeptide sequences of the present invention that maintain linear (in case of peptides) and conformational epitopes (in case of polypeptides) recognized by antibodies in the sera from individuals infected with HCV. It is within the scope of the invention to use for instance single or specific oligomeric antigens, dimeric antigens, as well as

combinations of single or specific oligomeric antigens. The HCVantigens of the present invention may be employed in virtually any assay format that employs a known antigen to detect antibodies. Of course, a format that denatures the HCV conformational epitope should be avoided or adapted. A common feature of all of these assays is that the antigen is contacted with the body component suspected of containing HCV antibodies under conditions that permit the antigen to bind to any such antibody present in the component. Such conditions will typically be physiologic temperature, pH and ionic strenght using an excess of antigen. The incubation of the antigen with the specimen is followed by detection of immune complexes comprised of the antigen.

Design of the immunoassays is subject to a great deal of variation, and many formats are known in the art. Protocols may, for example, use solid supports, or immunoprecipitation. Most assays involve the use of labeled antibody or polypeptide; the labels may be, for example, enzymatic, fluorescent, chemiluminescent, radioactive, or dye molecules. Assays which amplify the signals from the immune complex are also known; examples of which are assays which utilize biotin and avidin or streptavidin, and enzyme-labeled and mediated immunoassays, such as ELISA assays.

The immunoassay may be, without limitation, in a heterogeneous or in a homogeneous format, and of a standard or competitive type. In a heterogeneous format, the polypeptide is typically bound to a solid matrix or support to facilitate separation of the sample from the polypeptide after incubation. Examples of solid supports that can be used are nitrocellulose (e.g., in membrane or microtiter well form), polyvinyl chloride (e.g., in sheets or microtiter wells), polystyrene latex (e.g., in beads or microtiter plates, polyvinylidine fluoride (known as Immunolon™), diazotized paper, nylon membranes, activated beads, and Protein A beads. For example, Dynatech Immunolon™ 1 or Immunlon™ 2 microtiter plates or 0.25 inch polystyrene beads (Precision Plastic Ball) can be used in the heterogeneous format. The solid support containing the antigenic polypeptides is typically washed after separating it from the test sample, and prior to detection of bound antibodies. Both standard and competitive formats are know in the art.

In a homogeneous format, the test sample is incubated with the combination of antigens in solution. For example, it may be under conditions that will precipitate

any antigen-antibody complexes which are formed. Both standard and competitive formats for these assays are known in the art.

In a standard format, the amount of HCV antibodies in the antibody-antigen complexes is directly monitored. This may be accomplished by determining whether labeled anti-xenogeneic (e.g. anti-human) antibodies which recognize an epitope on anti-HCV antibodies will bind due to complex formation. In a competitive format, the amount of HCV antibodies in the sample is deduced by monitoring the competitive effect on the binding of a known amount of labeled antibody (or other competing ligand) in the complex.

Complexes formed comprising anti-HCV antibody (or in the case of competitive assays, the amount of competing antibody) are detected by any of a number of known techniques, depending on the format. For example, unlabeled HCV antibodies in the complex may be detected using a conjugate of anti-xenogeneic Ig complexed with a label (e.g. an enzyme label).

In an immunoprecipitation or agglutination assay format the reaction between the HCV antigens and the antibody forms a network that precipitates from the solution or suspension and forms a visible layer or film of precipitate. If no anti-HCV antibody is present in the test specimen, no visible precipitate is formed.

There currently exist three specific types of particle agglutination (PA) assays. These assays are used for the detection of antibodies to various antigens when coated to a support. One type of this assay is the hemagglutination assay using red blood cells (RBCs) that are sensitized by passively adsorbing antigen (or antibody) to the RBC. The addition of specific antigen antibodies present in the body component, if any, causes the RBCs coated with the purified antigen to agglutinate.

To eliminate potential non-specific reactions in the hemagglutination assay, two artificial carriers may be used instead of RBC in the PA. The most common of these are latex particles. However, gelatin particles may also be used. The assays utilizing either of these carriers are based on passive agglutination of the particles coated with purified antigens.

The HCV antigens of the present invention comprised of conformational epitopes will typically be packaged in the form of a kit for use in these immunoassays. The kit will normally contain in separate containers the native HCV antigen, control antibody formulations (positive and/or negative), labeled antibody

when the assay format requires the same and signal generating reagents (e.g. enzyme substrate) if the label does not generate a signal directly. The native HCV antigen may be already bound to a solid matrix or separate with reagents for binding it to the matrix. Instructions (e.g. written, tape, CD-ROM, etc.) for carrying out the assay usually will be included in the kit.

Immunoassays that utilize the native HCV antigen are useful in screening blood for the preparation of a supply from which potentially infective HCV is lacking. The method for the preparation of the blood supply comprises the following steps. Reacting a body component, preferably blood or a blood component, from the individual donating blood with HCV polypeptides of the present invention to allow an immunological reaction between HCV antibodies, if any, and the HCV antigen. Detecting whether anti-HCV antibody - HCV antigen complexes are formed as a result of the reacting. Blood contributed to the blood supply is from donors that do not exhibit antibodies to the native HCV antigens.

In cases of a positive reactivity to the HCV antigen, it is preferable to repeat the immunoassay to lessen the possibility of false positives. For example, in the large scale screening of blood for the production of blood products (e.g. blood transfusion, plasma, Factor VIII, immunoglobulin, etc.) 'screening' tests are typically formatted to increase sensitivity (to insure no contaminated blood passes) at the expense of specificity; i.e. the false-positive rate is increased. Thus, it is typical to only defer for further testing those donors who are 'repeatedly reactive'; i.e. positive in two or more runs of the immunoassay on the donated sample. However, for confirmation of HCV-positivity, the 'confirmation' tests are typically formatted to increase specificity (to insure that no false-positive samples are confirmed) at the expense of sensitivity.

The solid phase selected can include polymeric or glass beads, nitrocellulose, microparticles, microwells of a reaction tray, test tubes and magnetic beads. The signal generating compound can include an enzyme, a luminescent compound, a chromogen, a radioactive element and a chemiluminescent compound. Examples of enzymes include alkaline phosphatase, horseradish peroxidase and beta-galactosidase. Examples of enhancer compounds include biotin, anti-biotin and avidin. Examples of enhancer compounds binding members include biotin, anti-biotin and avidin. In order to block the effects of rheumatoid factor-like substances, the

test sample is subjected to conditions sufficient to block the effect of rheumatoid factor-like substances. These conditions comprise contacting the test sample with a quantity of anti-human IgG to form a mixture, and incubating the mixture for a time and under conditions sufficient to form a reaction mixture product substantially free of rheumatoid factor-like substance.

The present invention particularly relates to an immunoassay format in which the polypeptides (or peptides) of the invention are coupled to a membrane in the form of parallel lines . This assay format is particularly advantageous for HCV typing purposes.

The present invention also relates to a pharmaceutical composition comprising at least one (recombinant) polypeptides as defined above and a suitable excipient, diluent or carrier.

The present invention also relates to a method of preventing HCV infection, comprising administering the pharmaceutical composition as defined above to a mammal in effective amount to stimulate the production of protective antibody or protective T-cell response.

The present invention relates to the use of a composition as defined above in a method for preventing HCV infection.

The present invention further relates to a vaccine for immunizing a mammal against HCV infection, comprising at least one (recombinant) polypeptide as defined above, in a pharmaceutically acceptable carrier.

The term 'immunogenic' refers to the ability of a substance to cause a humoral and/or cellular response, whether alone or when linked to a carrier, in the presence or absence of an adjuvant. 'Neutralization' refers to an immune response that blocks the infectivity, either partially or fully, of an infectious agent. A 'vaccine' is an immunogenic composition capable of eliciting protection against HCV, whether partial or complete. A vaccine may also be useful for treatment of an individual, in which case it is called a therapeutic vaccine.

The term 'therapeutic' refers to a composition capable of treating HCV infection. The term 'effective amount' refers to an amount of epitope-bearing polypeptide sufficient to induce an immunogenic response in the individual to which it is administered, or to otherwise detectably immunoreact in its intended system (e.g., immunoassay). Preferably, the effective amount is sufficient to effect

treatment, as defined above. The exact amount necessary will vary according to the application. For vaccine applications or for the generation of polyclonal antiserum / antibodies, for example, the effective amount may vary depending on the species, age, and general condition of the individual, the severity of the condition being treated, the particular polypeptide selected and its mode of administration, etc. It is also believed that effective amounts will be found within a relatively large, non-critical range. An appropriate effective amount can be readily determined using only routine experimentation. Preferred ranges of proteins for prophylaxis of HCV disease are 0.01 to 100 μg/dose, preferably 0.1 to 50 μg/dose. Several doses may be needed per individual in order to achieve a sufficient immune response and subsequent protection against HCV disease.

The present invention also relates to a vaccine as defined above, comprising at least one (recombinant) polypeptide as defined above, with said polypeptide being unique for at least one of the subtypes or types as defined above.

Said vaccine compositions may include prophylactic as well as therapeutic vaccine compositions.

Pharmaceutically acceptable carriers include any carrier that does not itself induce the production of antibodies harmful to the individual receiving the composition. Suitable carriers are typically large, slowly metabolized macromolecules such as proteins, polysaccharides, polylactic acids, polyglycolic acids, polymeric amino acids, amino acid copolymers; and inactive virus particles. Such carriers are well known to those of ordinary skill in the art.

Preferred adjuvants to enhance effectiveness of the composition include, but are not limited to : aluminim hydroxide (alum), N-acetyl-muramyl-L-threonyl-D-isoglutamine (thr-MDP) as found in U.S. Patent No. 4,606,918, N-acetyl-normuramyl-L-alanyl-D-isoglutamine (nor-MDP), N-acetylmuramyl-L-alanyl-D-isoglutaminyl-L-alanine-2-(1'-2'-dipalmitoyl-sn-glycero-3-hydroxyphosphoryloxy)-ethylamine (MTP-PE) and RIBI, which contains three components extracted from bacteria, monophosphoryl lipid A, trehalose dimycolate, and cell wall skeleton (MPL + TDM + CWS) in a 2% squalene/Tween 80 emulsion. Any of the 3 components MPL, TDM or CWS may also be used alone or combined 2 by 2. Additionally, adjuvants such as Stimulon (Cambridge Bioscience, Worcester, MA)

Immunogenic compositions used as vaccines comprise a 'sufficient amount' or 'an immunologically effective amount' of the proteins of the present invention, as well as any other of the above mentioned components, as needed. 'Immunologically effective amount', means that the administration of that amount to an individual, either in a single dose or as part of a series, is effective for treatment, as defined above. This amount varies depending upon the health and physical condition of the individual to be treated, the taxonomic group of individual to be treated (e.g. nonhuman primate, primate, etc.), the capacity of the individual's immune system to synthesize antibodies, the degree of protection desired, the formulation of the vaccine, the treating doctor's assessment of the medical situation, the strain of infecting HCV, and other relevant factors. It is expected that the amount will fall in a relatively broad range that can be determined through routine trials. Usually, the amount will vary from 0.01 to 1000 μg/dose, more particularly from 0.1 to 100 μg/dose.

The proteins of the invention may also serve as vaccine carriers to present homologous (e.g. T cell epitopes or B cell epitopes fromfor istance the core,E1, E2, NS2, NS3, NS4 or NS5 regions) or heterologous (non-HCV) haptens, in the same manner as Hepatitis B surface antigen (see European Patent Application 174,444). In this use, envelope proteins provide an immunogenic carrier capable of stimulating an immune response to haptens or antigens conjugated to the aggregate. The antigen may be conjugated either by conventional chemical methods, or may be cloned into the gene encoding E1 and/or E2 at a location corresponding to a hydrophilic region

of the protein. Such hydrophylic regions include the V1 region (encompassing amino acid positions 191 to 202), the V2 region (encompassing amino acid positions 213 to 223), the V3 region (encompassing amino acid positions 230 to 242), the V4 region (encompassing amino acid positions 230 to 242), the V5 region (encompassing amino acid positions 294 to 303) and the V6 region (encompassing amino acid positions 329 to 336). Another useful location for insertion of haptens is the hydrophobic region (encompassing approximately amino acid positions 264 to 293). It is shown in the present invention that this region can be deleted without affecting the reactivity of the deleted E1 protein with antisera. Therefore, haptens may be inserted at the site of the deletion.

The immunogenic compositions are conventionally administered parenterally, typically by injection, for example, subcutaneously or intramuscularly. Additional formulations suitable for other methods of administration include oral formulations and suppositories. Dosage treatment may be a single dose schedule or a multiple dose schedule. The vaccine may be administered in conjunction with other immunoregulatory agents.

The administration of the immunogen(s) of the present invention may be for either a prophylactic or therapeutic purpose. When provided prophylactically, the immunogen(s) is provided in advance of any exposure to HCV or in advance of any symptom of any symptoms due to HCV infection. The prophylactic administration of the immunogen serves to prevent or attenuate any subsequent infection of HCV in a mammal. When provided therapeutically, the immunogen(s) is provided at (or shortly after) the onset of the infection or at the onset of any symptom of infection or disease caused by HCV. The therapeutic administration of the immunogen(s) serves to attenuate the infection or disease.

In addition to use as a vaccine, the compositions can be used to prepare antibodies to HCV (E1) proteins. The antibodies can be used directly as antiviral agents. To prepare antibodies, a host animal is immunized using the E1 proteins native to the virus particle bound to a carrier as described above for vaccines. The host serum or plasma is collected following an appropriate time interval to provide a composition comprising antibodies reactive with the (E1) protein of the virus particle. The gamma globulin fraction or the IgG antibodies can be obtained, for example, by use of saturated ammonium sulfate or DEAE Sephadex, or other

techniques known to those skilled in the art. The antibodies are substantially free of many of the adverse side effects which may be associated with other anti-viral agents such as drugs.

The present invention also relates particularly to a peptide corresponding to an amino acid sequence encoded by at least one of the HCV genomic sequences as defined above, with said peptide being unique to any of the HCV subtypes or types as defined in Table 5, and which contains at least one amino acid differing from any of the known HCV types or subtypes, or an analog thereof being substantially homologous and biologically equivalent.

The present invention relates particularly to a peptide comprising at least one unique epitope of the new sequences of the invention as represented in SEQ ID NO 1 to 106.

The present invention relates also particularly to a peptide comprising in its sequence a unique amino acid residue of the invention as defined above.

The present invention relates particularly to a peptide which is biotinylated as explained in WO 93/18054.

All the embodiments (immunoassay formats, vaccines, compositions, uses, etc.) illustrated for the polypeptides of the invention as above also relate to the peptides of the invention.

The present invention also relates to a method for detecting antibodies to HCV present in a biological sample, comprising:

(i) contacting the biological sample to be analysed for the presence of HCV with a peptide as defined above,

(ii) detecting the immunological ccomplex formed between said antibodies and said peptide.

The present invention also relates to a method for HCV typing, comprising:

(i) contacting the biological sample to be analysed for the presence of HCV with a peptide as defined above,

(ii) detecting the immunological ccomplex formed between said antibodies and said peptide.

The present invention also relates to a diagnostic kit for use in detecting the presence of HCV, said kit comprising at least one peptide as defined above, with said peptide being preferably bound to a solid support.

The present invention also relates to a diagnostic kit for HCV typing, said kit comprising at least one peptide as defined above, with said peptide being preferably bound to a solid support.

The present invention also relates to a diagnostic kit as defined above, wherein said peptides are selected from the following:

- at least one NS4 peptide,
- at least one NS4 peptide and at least one Core peptide,
- at least one NS4 peptide and at least one Core peptide and at least one E1 peptide,
- at least one NS4 peptide and at least one E1 peptide.

The present invention also relates to a diagnostic kit as defined above, said kit comprising a range of said peptides which are attached to specific locations on a solid substrate.

The present invention also relates to a diagnostic kit as defined above, wherein said solid support is a membrane strip and said peptides are coupled to the membrane in the form of parallel lines.

The present invention also relates to a pharmaceutical composition comprising at least one as defined above and a suitable excipient, diluent or carrier.

the present invention also relates to a method of preventing HCV infection, comprising administering the pharmaceutical composition as defined above to a mammal in effective amount to stimulate the production of protective antibody or protective T-cell response.

The present invention also relates to the use of a composition as defined above in a method for preventing HCV infection.

The present invention also relates to a vaccine for immunizing a mammal against HCV infection, comprising at least one peptide as defined above, in a pharmaceutically acceptable carrier.

The present invention relates also to a vaccine as defined above, comprising at least one peptide as defined above, with said peptide being unique for at least one of the subtypes or types as defined in Table 5.

The present invention relates to an antibody raised upon immunization with at least one polypeptide or peptide as defined above, with said antibody being specifically reactive with any of said polypeptides or peptides, and with said antibody being preferably a monoclonal antibody.

The monoclonal antibodies of the invention can be produced by any hybridoma liable to be formed according to classical methods from splenic cells of an animal, particularly from a mouse or rat, immunized against the HCV polypeptides according to the invention as defined above on the one hand, and of cells of a myeloma cell line on the other hand, and to be selected by the ability of the hybridoma to produce the monoclonal antibodies recognizing the polypeptides which has been initially used for the immunization of the animals.

The antibodies involved in the invention can be labelled by an appropriate label of the enzymatic, fluorescent, or radioactive type.

The monoclonal antibodies according to this preferred embodiment of the invention may be humanized versions of mouse monoclonal antibodies made by means of recombinant DNA technology, departing from parts of mouse and/or human genomic DNA sequences coding for H and L chains or from cDNA clones coding for H and L chains.

Alternatively the monoclonal antibodies according to this preferred embodiment of the invention may be human monoclonal antibodies. These antibodies according to the present embodiment of the invention can also be derived from human peripheral blood lymphocytes of patients infected with HCV type 1 subtype 1d, 1e, 1f or 1g, HCV type 2 subtype 2e, 2f, 2g, 2h, 2i, 2k or 2l; HCV type 3, subtype 3g; HCV type 4 subtype 4k, 4l or 4m; and/or HCV type 7 (subtypes 7a, 7c or 7d), 9, 10 or 11, or vaccinated against HCV. Such human monoclonal antibodies are prepared, for instance, by means of human peripheral blood lymphocytes (PBL) repopulation of severe combined immune deficiency (SCID) mice (for recent review, see Duchosal et al. 1992) or by screening Eppstein Barr-virus-transformed lymphocytes of infected or vaccinated individuals for the presence of reactive B-cells by means of the antigens of the present invention.

The invention also relates to the use of the proteins of the invention, muteins thereof, or peptides derived therefrom for the selection of recombinant antibodies by the process of repertoire cloning (Persson et al., 1991).

Antibodies directed to peptides derived from a certain genotype may be used either for the detection of such HCV genotypes, or as therapeutic agents.

The present invention relates also to a method for detecting HCV antigens present in a biological sample, comprising:

(i) contacting said biological sample with an antibody as defined above,

(ii) detecting the immune compleexes formed between said HCV antigens and said antibody.

The present invention relates also to a method for HCV typing, comprising:

(i) contacting said biological sample with an antibody as defined above,

(ii) detecting the immune compleexes formed between said HCV antigens and said antibody.

The present invention relates also to a diagnostic kit for use in detecting the presence of HCV, said kit comprising at least one antibody as defined above, with said antibody being preferably bound to a solid support.

The present invention relates also to a diagnostic kit for HCV typing, said kit comprising at least one antibody as defined above, with said antibody being preferably bound to a solid support.

The present invention relates also to a diagnostic kit as defined above, said kit comprising a range of said antibodies which are attached to specific locations on a solid substrate.

The present invention relates also to a pharmaceutical composition comprising at least one antibody as defined above and a suitable excipient, diluent or carrier.

The present invention relates also to a method of preventing or treating HCV infection, comprising administering the pharmaceutical composition as defined above to a mammal in effective amount.

The present invention relates also to the use of a composition as defined above in a method for preventing or treating HCV infection.

The genotype may also be detected by means of a type-specific antibody as defined above, which may also linked to any polynucleotide sequence that can afterwards be amplified by PCR to detect the immune complex formed (Immuno-PCR, Sano et al., 1992).

Any publications or patent applications referred to herein are incorporated by reference. The following examples illustrate aspects of the invention but are in no way intended to limit the scope thereof.

FIGURE LEGENDS

Figure Legends

Figure 1

Alignment of the nucleotide sequences of the Core/E1 region of some of the isolates of the newly identified types and subtypes of the present invention, with other known prototype isolates of subtypes.

Figure 2

Alignment of the amino acid sequences of the Core/E1 region of some of the isolates of the newly identified types and subtypes of the present invention, with other known prototype isolates of subtypes.

Figure 3

Nucleotide and amino acid sequences obtained from the new HCV isolates of the present invention (SEQ ID NO 1 to 106).

Figure 4

Alignment of the amino acid sequences of the Core/E1 region of some of the isolates of the newly identified types and subtypes of the present invention, with other known prototype isolates of subtypes.

Figure 5

Alignment of the nucleotide sequences of the NS5b region of some of the isolates of the newly identified types and subtypes of the present invention, with other known prototype isolates of subtypes.

Figure 6

Alignment of the amino acid sequences of the NS5b region of some of the isolates of the newly identified types and subtypes of the present invention, with other known prototype isolates of subtypes.

Table 5

Overview of the new subtypes and types of the present invention and the regions sequenced. The subtypes between barckets have been replaced by the non-bracketed subtypes following the classification of Tokita et al. (1994).

Examples

Serum samples.

Serum samples from Cameroonian blood donors (CAM) were screened for HCV antibodies with Innotest HCV Ab III, and confirmed by INNO-LIA HCV III (Innogenetics, Antwerp, Belgium). Serum samples from patients with chronic hepatitis C infection were obtained from various centers in the Benelux countries (BNL), from France (FR), from Pakistan (PAK), from Egypt (EG), and from Vietnam (VN).

Samples from the Benelux, Cameroon, France and Vietnam were selected because of their aberrant reactivities (isolates CAM1078, FR2, FR1, VN4, VN12, VN13, NE98 and others (see Table 5)).

cPCR, LiPA, cloning and sequencing.

RNA isolation, cDNA synthesis, PCR, cloning, and LiPA genotyping using biotinylated 5' UR amplification products were performed as described (Stuyver et al., 1994c). The 5' UR, the Core/E1, and the NS5B PCR products were used for direct sequencing. The sequence of the universal 5' UR primers HCPr95, HCPr96, HCPr98, and HCPr29, were described previously (Stuyver et al. 1993b). The following primers were also described (Stuyver et al. 1994c): HCPr41, a sense

primer for the amplification of the Core region; HCPr52 and HCPr54 for amplification of the Core/E1 region; and HCPr206 and HCPr207 for amplication of a 340-bp NS5B region.

Serum samples BNL1, BNL2, BNL3, BNL4, BNL5, BNL6, BNL7, BNL8, BNL9, BNL10, BNL11, BNL12, CAM1078, FR2, FR16, FR4, FR13, VN13, VN4, VN12, FR1, NE98, and FR19 were analyzed in the Core/E1 region by direct sequencing. Serum samples BNL1, BNL2, FR17, CAM1078, FR2, FR16, BNL3, FR4, BNL5, FR13, FR18, PAK64, BNL8, BNL12, EG81, VN13, VN4, VN12, FR1, NE98, FR14, FR15, and FR19 were also analyzed in the NS5B region by direct sequencing. Partial 5' UR, Core, E1, and NS5B sequences were obtained. The length of the obtained sequences is sufficient to classify the obtained sequences into new types or subtypes, based on the phylogenetic distances to known sequences. The following sequences could be obtained (nucleotide sequences have odd-numbered SEQ ID NO., amino acid sequences have even-numbered SEQ ID NO.): SEQ ID NO 1, 3, 5, 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103 and 105. The amino acid sequences deduced therefrom are given in SEQ ID NO 2, 4, 7, 9, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104 and 106. Table 5 gives an overview of these sequences.

Table 5

Type	Isolate	Nucleotide sequence position		
1d	BNL1	1-310 (SEQ ID NO. 1)	478-925 (SEQ ID NO. 3)	7932-8271 (SEQ ID NO. 53)
1d	BNL2	1-310 (SEQ ID NO. 5)	478-925 (SEQ ID NO. 7)	7932-8271 (SEQ ID NO. 55)
1d	FR17			7932-8271(SEQ ID NO. 57)
1e	CAM1078	1-223 (SEQ ID NO. 9)	(-238)-414 (SEQ ID NO. 59)	7932-8271 (SEQ ID NO. 61)
1f	FR2	1-950 (SEQ ID NO. 11)		7932-8271 (SEQ ID NO. 63)
1g	FR16	(-15)-816 (SEQ ID NO. 65)		7932-8271(SEQ ID NO. 67)
2e	BNL3	1-310 (SEQ ID NO. 13)	478-957 (SEQ ID NO. 15)	7932-8271(SEQ ID NO. 69)
2f	FR4	1-957 (SEQ ID NO 17)		7932-8271(SEQ ID NO. 71)
2g	BNL4		478-925 (SEQ ID NO. 19)	
2h	BNL5	1-310 (SEQ ID NO. 21)	478-925 (SEQ ID NO. 23)	7932-8271(SEQ ID NO. 73)
2i	BNL6		478-833 (SEQ ID NO. 25)	
2k	FR13	(-238)-957 (SEQ ID NO. 75)		7932-8271(SEQ ID NO. 77)
2l	FR18			7932-8271(SEQ ID NO. 79)
3g	PAK64			7932-8271(SEQ ID NO. 81)
4k	BNL7	1-310 (SEQ ID NO. 27)	478-925 (SEQ ID NO. 29)	
4k	BNL8		478-925 (SEQ ID NO. 31)	7932-8271 (SEQ ID NO. 83)
4k	BNL9		478-925 (SEQ ID NO. 33)	
4k	BNL10		478-925 (SEQ ID NO. 35)	
4k	BNL11		478-925 (SEQ ID NO. 37)	
4l	BNL12		478-925 (SEQ ID NO. 39)	7932-8271 (SEQ ID NO. 85)
4m	EG81			7932-8271 (SEQ ID NO. 87)
7a(8b)	VN13	1-413 (SEQ ID NO. 45)		7932-8271 (SEQ ID NO. 89)
7c(8a)	VN4	1-957 (SEQ ID NO. 43)		7932-8271 (SEQ ID NO. 91)
7d(9a)	VN12	1-957 (SEQ ID NO. 47)		7932-8271 (SEQ ID NO. 93)
9a(7a)	FR1	1-957 (SEQ ID NO. 41)		7932-8271 (SEQ ID NO. 95)
10a	NE98	1-310 (SEQ ID NO. 49)	478-925 (SEQ ID NO. 51)	7932-8271 (SEQ ID NO. 97)
11a	FR14			7932-8266 (SEQ ID NO. 99)
11a	FR15			7932-8271(SEQ ID NO. 101)
11a	FR19	(-238)-223 (SEQ ID NO. 103)		7932-8271 (SEQ ID NO. 105)

Table 5-continued

Type	Isolate	Amino acid sequence position		
1d	BNL1	1-103 (SEQ ID NO. 2)	159-308 (SEQ ID NO. 4)	2645-2757 (SEQ ID NO. 54)
1d	BNL2	1-103 (SEQ ID NO. 6)	159-308 (SEQ ID NO. 8)	2645-2757 (SEQ ID NO. 56)
1d	FR17			2645-2757 (SEQ ID NO. 58)
1e	CAM1078	1-74 (SEQ ID NO. 10)	1-138 (SEQ ID NO. 60)	2645-2757 (SEQ ID NO. 62)
1f	FR2	1-316 (SEQ ID NO. 12)		2645-2757 (SEQ ID NO. 64)
1g	FR16	1-158 (SEQ ID NO. 66)		2645-2757 (SEQ ID NO. 68)
2e	BNL3	1-103 (SEQ ID NO. 14)	159-317 (SEQ ID NO. 16)	2645-2757 (SEQ ID NO. 70)
2f	FR4	1-317 (SEQ ID NO. 18)		2645-2757 (SEQ ID NO. 72)
2g	BNL4		159-308 (SEQ ID NO. 20)	
2h	BNL5	1-103 (SEQ ID NO. 22)	159-308 (SEQ ID NO. 24)	2645-2757 (SEQ ID NO. 74)
2i	BNL6		159-277 (SEQ ID NO. 26)	
2k	FR13	1-316 (SEQ ID NO. 76)		2645-2757 (SEQ ID NO. 78)
2l	FR18			2645-2757 (SEQ ID NO. 80)
3g	PAK64			2645-2757 (SEQ ID NO. 82)
4k	BNL7	1-103 (SEQ ID NO. 28)	159-308 (SEQ ID NO. 30)	
4k	BNL8		159-308 (SEQ ID NO. 32)	2645-2757 (SEQ ID NO. 84)
4k	BNL9		159-308 (SEQ ID NO. 34)	
4k	BNL10		159-308 (SEQ ID NO. 36)	
4k	BNL11		159-308 (SEQ ID NO. 38)	
4l	BNL12		159-308 (SEQ ID NO. 40)	
4m	EG81			2645-2757 (SEQ ID NO. 86)
7a (8b)	VN13	1-137 (SEQ ID NO. 46)		2645-2757 (SEQ ID NO. 88)
7c (8a)	VN4	1-317 (SEQ ID NO. 44)		2645-2757 (SEQ ID NO. 90)
7d (9a)	VN12	1-317 (SEQ ID NO. 48)		2645-2757 (SEQ ID NO. 92)
9a (7a)	FR1	1-317 (SEQ ID NO. 42)		2645-2757 (SEQ ID NO. 94)
10a	NE98	1-103 (SEQ ID NO. 50)	159-308 (SEQ ID NO. 52)	2645-2757 (SEQ ID NO. 96)
11a	FR14			2645-2757 (SEQ ID NO. 98)
11a	FR15			2645-2755 (SEQ ID NO. 100)
11a	FR19	1-74 (SEQ ID NO. 104)		2645-2757 (SEQ ID NO. 102)
				2645-2757 (SEQ ID NO. 106)

Phylogenetic analysis.

Previously published sequences were taken from the EMBL/Genbank database. Alignments were created using the program HCVALIGN (Stuyver et al. 1994c). Sequences were presented in a sequential format to the Phylogeny Inference Package (PHYLIP) version 3.5c (public domain program freely available from the University of Washington, Seattle, USA). Distance matrices were produced by DNADIST using the Kimura 2-parameter setting and further analyzed in NEIGHBOR, using the neighbor-joining setting. The program DRAWTREE was used to create graphic outputs.

Identification of new subtypes

These analyses indicated the clustering of BNL1, BNL2, CAM 1078, FR2, FR16, and FR17 with type 1 isolates, yet neither of these sequences clustered together with any of the known type 1 subtypes 1a, 1b, or 1c. BNL1, BNL2, and FR17 clearly clustered together and could be assigned a new type 1 subtype 1d, while CAM1078 could be classified into another new subtype 1e, FR2 could be classified into another type 1 subtype 1f, and FR16 could be classified into yet another type 1 subtype 1g. Interestingly, all 3 type 1d isolates (BNL1, BNL2, and FR17) and 1g isolate FR16 were obtained from patients of Moroccan ethnic origin who resided in Europe.

Another group of isolates showed homology to other type 2 sequences, but none of the isolates BNL3, FR4, BNL4, BNL5, BNL6, FR13, or FR18 could be classified into one of the known type 2 subtypes 2a, 2b, 2c (Bukh et al., 1993), or 2d (Stuyver et al., 1994c). Based on the phylogenetic distances to other type 2 isolates and to other isolates of the group, each of these isolates could be classified into a new type 2 subtype. BNL3 was assigned subtype 2e, FR4 subtype 2f, BNL4 subtype 2g, BNL5 subtype 2h, and BNL6 could be classified into yet another type 2 subtype 2i. If the previously published isolate HN4 is classified as 2j, FR13 and FR18 may be classified into new type 2 subtypes 2k and 2l. However, the possibility that FR13 and FR18 could belong to subtypes 2g or 2i has not yet been ruled out. Definite classification can be obtained by determining the NS5B sequences of isolates BNL4 and BNL6, belonging to subtypes 2g and 2i, respectively.

Isolate PAK64 showed homology to type 3 sequences, but could not be classified into one of the known type 3 subtypes 3a to f. Based on the phylogenetic distances to other type 3 isolates, PAK64 could be classified into a new type 3

subtype. PAK64 was assigned subtype 3g. However, the possibility that PAK64 belongs to a known type 3 subtype can not be strictly ruled out since only one region of the genome has been sequenced. Definite classification can be obtained by determining the Core/E1 sequences of isolate PAK64 after amplification with primerHcPr52 and HcPr54.

Among the Benelux and Egyptian samples that were analyzed, some sequences clustered with the previously identified type 4 subtypes 4c and 4d. However, BNL7, BNL8, BNL9, BNL10, BNL11, BNL12, and EG81 clustered into new subtypes of type 4. Isolates BNL7, BNL8, BNL9, BNL10, and BNL11 clustered again separately from BNL12 and EG81 into a new subtype 4k. This subtype was the predominant subtype in the Benelux countries. BNL12 and EG81 also segregated into separate subtypes. BNL12 was assigned to another new subtype 4l and EG81 was assigned to yet another new subtype 4m.

Identification of new HCV major types

Isolates FR1, VN4, VN12, VN13, NE98, FR14, FR15, and FR19 did not cluster with any of the known 6 major types of HCV. VN4, VN12, and VN13 were very distantly related to genotype 6, but phylogenetic analysis indicated that these isolates should be assigned new major types. VN13, VN4 and VN12 were related at the subtype level and assigned type 7a, 7c, and 7d, respectively. FR1 was not related to any known isolate and was assigned genotype 9a. NE98 shows a distant relatedness to type 3 sequences, yet phylogenetic analysis suggested classification into a new major type 10a. Depending on international guidelines for assigning type and subtype levels, NE98 may also be classified into an additional type 3 subtype. FR14, FR15, and FR19 show a very distant relatedness to type 2 sequences, yet phylogenetic analysis indicated thes isolates to be classified into a new major type 11, all belonging to the same subtype designated 11a. Depending on international guidelines for assigning type and subtype levels, FR14, FR15, and FR19 may also be classified into an additional type 2 subtype.

REFERENCES

Barany F (1991). Genetic disease detection and DNA amplification using cloned thermostable ligase. Proc Natl Acad Sci USA 88: 189-193.

Bej A, Mahbubani M, Miller R, Di Cesare J, Haff L, Atlas R (1990) Mutiplex PCR amplification and immobilized capture probes for detection of bacterial pathogens and indicators in water. Mol Cell Probes 4:353-365.

Bukh J, Purcell R, Miller R (1992). Sequence analysis of the 5' noncoding region of hepatitis C virus. Proc Natl Acad Sci USA 89:4942-4946.

Bukh J, Purcell R, Miller R (1993). At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide. Proc. Natl. Acad. Sci. USA 90,8234-8238.

Cha T, Beal E, Irvine B, Kolberg J, Chien D, Kuo G, Urdea M (1992) At least five related, but distinct, hepatitis C viral genotypes exist. Proc Natl Acad Sci USA 89:7144-7148.

Chan S-W, Simmonds P, McOmish F, Yap P, Mitchell R, Dow B, Follett E (1991) Serological responses to infection with three different types of hepatitis C virus. Lancet 338:1991.

Chan S-W, McOmish F, Holmes E, Dow B, Peutherer J, Follett E, Yap P, Simmonds P (1992) Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants. J Gen Virol 73:1131-1141.

Chomczynski P, Sacchi N (1987) Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal Biochem 162:156-159.

Choo Q, Richman K, Han J, Berger K, Lee C, Dong C, Gallegos C, Coit D, Medina-

Selby A, Barr P, Weiner A, Bradley D, Kuo G, Houghton M (1991) Genetic organization and diversity of the hepatitis C virus. Proc Natl Acad Sci USA 88:2451-2455.

Compton J (1991). Nucleic acid sequence-based amplification. Nature, 350: 91-92.

Duchosal A, Eming S, Fisher P (1992) Immunization of hu-PBL-SCID mice and the resue of human monoclonal Fab fragments through combinatorial libraries. Nature 355:258-262.

Duck P (1990). Probe amplifier system based on chimeric cycling oligonucleotides. Biotechniques 9, 142-147.

Guatelli J, Whitfield K, Kwoh D, Barringer K, Richman D, Gengeras T (1990) Isothermal, in vitro amplification of nucleic acids by a multienzyme reaction modeled after retroviral replication. Proc Natl Acad Sci USA 87: 1874-1878.

Hijikata M, Kato N, Ootsuyama Y, Nakagawa M, Shimotohmo K (1991) Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis. Proc Natl Acad Sci USA 88, 5547-5551.

Jacobs K, Rudersdorf R, Neill S, Dougherty J, Brown E, Fritsch E (1988) The thermal stability of oligonucleotide duplexes is sequence independent in tetraalkylammonium salt solutions: application to identifying recombinant DNA clones. Nucl Acids Res 16:4637-4650.

Kato N, Hijikata M, Ootsuyama Y, Nakagawa M, Ohkoshi S, Sugimura T, Shimotohno K (1990) Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. Proc Natl Acad Sci USA 87:9524-9528.

Kwoh D, Davis G, Whitfield K, Chappelle H, Dimichele L, Gingeras T (1989). Transcription-based amplification system and detection of amplified human immunodeficiency virus type 1 with a bead-based sandwich hybridization format. Proc Natl Acad Sci USA, 86: 1173-1177.

Kwok S, Kellogg D, McKinney N, Spasic D, Goda L, Levenson C, Sinisky J, (1990). Effects of primer-template mismatches on the polymerase chain reaction: Human immunodeficiency views type 1 model studies. Nucl. Acids Res., 18: 999.

Landgren U, Kaiser R, Sanders J, Hood L (1988). A ligase-mediated gene detection technique. Science 241:1077-1080.

Lizardi P, Guerra C, Lomeli H, Tussie-Luna I, Kramer F (1988) Exponential amplification of recombinant RNA hybridization probes. Bio/Technology 6:1197-1202.

Lomeli H, Tyagi S, Printchard C, Lisardi P, Kramer F (1989) Quantitative assays based on the use of replicatable hybridization probes. Clin Chem 35: 1826-1831.

Machida A, Ohnuma H, Tsuda F, Munekata E, Tanaka T, Akahane Y, Okamoto H, Mishiro S (1992) Hepatology 16, 886-891.

Maniatis T, Fritsch E, Sambrook J (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Mori S, Kato N, Yagyu A, Tanaka T, Ikeda Y, Petchclai B, Chiewsilp P, Kurimura T, Shimotohno K (1992) A new type of hepatitis C virus in patients in Thailand. Biochem Biophys Res Comm 183:334-342.

Okamoto H, Okada S, Sugiyama Y, Kurai K, Iizuka H, Machida A, Miyakawa Y, Mayumi M (1991) Nucleotide sequence of the genomic RNA of hepatitis C virus isolated from a human carrier: comparison with reported isolates for conserved and divergent regions. J Gen Virol 72:2697-2704.

Okamoto H, Kurai K, Okada S, Yamamoto K, Lizuka H, Tanaka T, Fukuda S, Tsuda F, Mishiro S (1992) Full-length sequences of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes. Virology 188:331-341.

Persson M, Caothien R, Burton D (1991). Generation of diverse high-affinity human monoclonal antibodies by repertoire cloning. Proc Natl Acad Sci USA 89:2432-2436.

Saiki R, Gelfand D, Stoffel S, Scharf S, Higuchi R, Horn G, Mullis K, Erlich H (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487-491.

Saiki R, Walsh P, Levenson C, Erlich H (1989) Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes (1989) Proc Natl Acad Sci USA 86:6230-6234.

Sano T, Smith C, Cantor C (1992) Immuno-PCR: very sensitive antigen detection by means of specific antibody-DNA conjugates. Science 258:120-122.

Simmonds P, McOmsh F, Yap P, Chan S, Lin C, Dusheiko G, Saeed A, Holmes E (1993a), Sequence variability in the 5' non-coding region of hepatitis C virus : identification of a new virus type and restrictions on sequence diversity. J Gen Virology, 74:661-668.

Stuyver L, Rossau R, Wyseur A, Duhamel M, Vanderborght B, Van Heuverswyn H, Maertens G (1993b) Typing of hepatitis C virus (HCV) isolates and characterization of new (sub)types using a Line Probe Assay. J Gen Virology, 74: 1093-1102.

Tokita et al. (1994) Hepatitis C virus vraiants from Vietnam are classifiable into the seventh, eighth, and ninth major genetic groups. Proc. Natl. Acad. Sci, 91: 11022-11026.

Walker G, Little M, Nadeau J, Shank D (1992). Isothermal in vitro amplification of DNA by a restriction enzyme/DNA polymerase system. Proc Natl Acad Sci USA 89:392-396.

Wu D, Wallace B (1989). The ligation amplification reaction (LAR) - amplification of specific DNA sequences using sequential rounds of template-dependent ligation. Genomics 4:560-569.

Miller P, Yano J, Yano E, Carroll C, Jayaram K, Ts'o P (1979) Nonionic nucleic acid analogues. Synthesis and characterization of dideoxyribonucleoside methylphosphonates. Biochemistry 18(23):5134-43.

Nielsen P, Egholm M, Berg R, Buchardt O (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. Science 254(5037):1497-500.

Nielsen P, Egholm M, Berg R, Buchardt O (1993) Sequence specific inhibition of DNA restriction enzyme cleavage by PNA. Nucleic-Acids-Res. 21(2):197-200.

Asseline U, Delarue M, Lancelot G, Toulme F, Thuong N (1984) Nucleic acid-binding molecules with high affinity and base sequence specificity : intercalating agents covalently linked to oligodeoxynucleotides. Proc. Natl. Acad. Sci. USA 81(11):3297-301.

Matsukura M, Shinozuka K, Zon G, Mitsuya H, Reitz M, Cohen J, Broder S (1987) Phosphorothioate analogs of oligodeoxynucleotides : inhibitors of replication and cytopathic effects of human immunodeficiency virus. Proc. Natl. Acad. Sci. USA 84(21):7706-10.

Maertens, G., Ducatteeuw, A., Stuyver, L., Vandeponseele, P., Venneman, A., Wyseur, A., Bosman, F., Heijtink, R. & de Martynoff, G. (1994) Low prevalence of anti-E1 antibodies reactive to recombinant type 1b E1 envelope protein in type 2, 3, and 4 sera, but high prevalence in subtypes 1a and 1b. In: Viral Hepatitis and Liver Disease,, Proceedings of the International Symposium on Viral Hepatitis and Liver Disease (Eds. Nishioka, K., Suzuki, H., Mishiro, S., and Oda, T.), pp 314-316, Springer-Verlag Tokyo.

Simmonds, P., Rose, K.A., Graham, S., Chan, S.-W., McOmish, F., Dow, B.C., Follett, E.A.C., Yap, P.L., & Marsden, H. (1993b) Mapping of serotype-specific, immunodominant epitopes in the NS4 region of hepatitis C virus (HCV): Use of type-specific peptides to serologically discriminate infections with HCV type 1, 2, and 3.

J. Clin. Microbiol. **31**, 1493-1503.

Simmonds, P., Holmes, E.C., Cha, T.-A., Chan, S.-W., McOmish, F., Irvine, B., Beall, E., Yap, P.L., Kolberg, J., & Urdea, M.S. (1993c) *J. Gen. Virol.* **74**, 2391-2399.

Stuyver, L., Van Arnhem, W., Wyseur, A. & Maertens, G. (1994) Cloning and phylogenetic analysis of the Core, E2, and NS3/4 regions of hepatitis C virus type 5a. Biochem. Biophys. Res. Comm. **202**, 1308-1314.

Simmonds, P., Alberti, A., Alter, H., Bonino, F., Bradley, D.W., Bréchot, C., Brouwer, J., Chan, S.-W., Chayama K., Chen, D.-S., Choo, Q.-L., Colombo, M., Cuypers, T., Date, T., Dusheiko, G., Esteban, J.I., Fay, O., Hadziyannis, S., Han, J., Hatzakis, A., Holmes, E.C., Hotta, H., Houghton, M., Irvine, B., Kohara, M., Kolberg, J.A., Kuo, G., Lau, J.Y.N., Lelie, P.N., Maertens, G., McOmish, F., Miyamura, T., Mizokami, M., Nomoto, A., Prince A.M., Reesink, H.W., Rice, C., Roggendorf, M., Schalm, S., Shikata, T., Shimotohno, K., Stuyver, L., Trépo, C., Weiner, A., Yap, P.L. & Urdea, M.S. (1994) A proposed system for the nomenclature of hepatitis C virus genotypes. *Hepatology* **19**, 1321-1324.

Stuyver, L., Van Arnhem, W., Wyseur, A., DeLeys, R. & Maertens, G. (1993a) Analysis of the putative E1 envelope and NS4a epitope regions of HCV type 3. Biochem. Biophys. Res. Comm. **192**, 635-641.

Stuyver, L., Rossau, R., Wyseur, A., Duhamel, M., Vanderborght, B., Van Heuverswyn, H. & Maertens, G. (1993b) Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. J. Gen Virol. **74**, 1093-1102.

Stuyver, L., Wyseur, A., Van Arnhem, W., Rossau, R., Delaporte, E., Dazza, M.-C., Van Doorn, L.-J., Kleter, B. & Maertens, G. (1994a) The use of a line probe assay as a tool to detect new types or subtypes of hepatitis C virus. In: Viral Hepatitis and Liver Disease, Proceedings of the International Symposium on Viral Hepatitis and Liver Disease (Eds. Nishioka, K., Suzuki, H., Mishiro, S., and Oda, T.), pp 317-319,

Springer-Verlag Tokyo.

Stuyver, L., Van Arnhem, W., Wyseur, A. & Maertens, G. (1994b) Cloning and Phylogenetic analysis of the Core, E2, and NS3/4 regions of the hepatitis C virus type 5a. Biochem. Biophys. Res. Comm. **202**, 1308-1314.

Stuyver, L., Van Arnhem, W., Wyseur, A., Hernandez, F., Delaporte, E., & Maertens, G. (1994c) Classification of hepatitis C viruses based on phylogenetics analysis of the E1 and NS5B regions and identification of 5 new subtypes. Proc. Natl. Acad. Sci. USA **91**.

Stuyver et al. (1995) Hepatitis C virus genotyping by means of 5'-UR/core line probe assays and molecular analysis of untypeable samples. Virus Reasearch (in press).

CLAIMS

1. An HCV polynucleic acid, having a nucleotide sequence which is unique to a theretofore unidentified HCV type or subtype which is different from HCV subtypes 1a, 1b, 1c, 2a, 2b, 2c, 2d, 3a, 3b, 3c, 3d, 3e, 3f, 4a, 4b, 4c, 4d, 4e, 4f, 4g, 4h, 4i, 4j, 5a or 6a, with said HCV subtypes being classified as in Table 3 by comparison of a part of the NS5 gene nucleotide sequence spanning positions 7932 to 8271, with said amino acid numbering being shown in Table 1, and with said polynucleic acid containing at least one nucleotide differing from said known HCV nucleotide sequences, or the complement thereof.

2. A polynucleic acid according to claim 1, having a nucleotide sequence which is unique to at least one of HCV subtypes 1d, 1e, 1f, 1g, 2e, 2f, 2g, 2h, 2i, 2k, 2l, 3g, 4k, 4l, 4m, 7a, 7c or 7d, with said HCV subtypes being classified as defined in claim 1.

3. A polynucleic acid according to claim 1, having a nucleotide sequence which is unique to at least one of HCV types 9, 10 or 11, with said HCV types being classified as defined in claim 1.

4. A polynucleic acid according to any of claims 1 to 3 encoding an HCV polyprotein comprising in its amino acid sequence at least one of the following amino acid residues:
I15, C38, V44, A49, Q43, P49, Q55, A58, S60 or D60, E68 or V68, H70, A71 or Q71 or N71, D72, H81, H101, D106, S110, L130, I134, E135, L140, S148, T150 or E150, Q153, F155, D157, G160, E165, I169, F181, L186, T190, T192 or I192 or H192, I193, A195, S196, R197 or N197 or K197, Q199 or D199 or H199 or N199, F200 or T200, A208, I213, M216 or S216, N217 or S217 or G217 or K217, T218, I219, A222, Y223, I230, W231 or L231, S232 or H232 or A232, Q233, E235 or L235, F236 or T236, F237, L240 or M240, A242, N244, N249, I250 or K250 or R250, A252 or C252, A254, I255 or V255, D256 or M256, E257, E260 or K260, R261, V268, S272 or R272, I285, G290 or F290, A291, A293 or L293 or W293, T294 or A294, S295 or H295, K296 or E296, Y297 or M297, I299 or Y299, I300,

S301, P316, S2646, A2648, G2649, A2650, V2652, Q2653, H2656 or L2656, D2657, F2659, K2663 or Q2663, A2667 or V1667, D2677, L2681, M2686 or Q2686 or E2686, A2692 or K2692, H2697, I2707, L2708 or Y2708, A2709, A2719 or M2719, F2727, T2728 or D2728, E2729, F2730 or Y2730, I2741, I2745, V2746 or E2746 or L2746 or K2746, A2748, S2749 or P2749, R2750, E2751, D2752 or N2752 or S2752 or T2752 or V2752 or I2752 or Q2752, S2753 or D2753 or G2753, D2754, A2755, L2756 or Q2756, R2757,

with said notation being composed of a letter representing the amino acid residue by its one-letter code, and a number representing the amino acid numbering as shown in Table 1,

or a part of said polynucleic acid which is unique to at least one of the HCV subtypes or types as defined in claims 2 to 3, and which contains at least one nucleotide differing from known HCV nucleotide sequences, or the complement thereof.

5. A polynucleic acid according to any of claims 1 to 4, with said polynucleic acid encoding a HCV polyprotein comprising in its amino acid sequence at least one amino acid sequence chosen from the following list:

ARQSDGRSWAQ or ARRSEGRSWAQ as for subtype 1d	(SEQ ID NO 107 and 108)
ERRPEGRSWAQ as for subtype 1e	(SEQ ID NO 109)
ARRPEGRSWAQ as for subtype 1f	(SEQ ID NO 110)
DRRTTGKSWGR as for subtype 2k	(SEQ ID NO 111)
DRRATGRSWGR as for subtype 2e	(SEQ ID NO 112)
DRRATGKSWGR as for subtype 2f	(SEQ ID NO 113)
VRQPTGRSWGQ as for type 9	(SEQ ID NO 114)
VRHQTGRTWAQ as for subtype 7a and 7c	(SEQ ID NO 115)
VRQNQGRTWAQ as for subtype 7d	(SEQ ID NO 116)
ARRTEGRSWAQ as for type 10	(SEQ ID NO 117)
VRRTTGRXXXX or VRRTTGRTWAQ as for type 11	(SEQ ID NO 118 and 119)
HEVRNASGVYHV or HEVRNASGVYHL as for subtype 1d	(SEQ ID NO 120 and 121)
YEVHSTTDGYHV as for subtype 1f	(SEQ ID NO 122)
VEVKNTSQAYMA as for subtype 2e	(SEQ ID NO 123)
IQVKNNSHFYMA as for subtype 2f	(SEQ ID NO 124)

VQVKNTSTMYMA as for subtype 2g (SEQ ID NO 125)

VQVKNTSHSYMV as for subtype 2h (SEQ ID NO 126)

VQVANRSGSYMV as for subtype 2i (SEQ ID NO 127)

VEIKNTXNTYVL or VEIKNTSNTYVL as for subtype 2k (SEQ ID NO 128 and 129)

INYRNVSGIYYV or INYRNTSGIYHV or INYHNTSGIYHI or TNYRNVSGIYHV as for subtype 4k (SEQ ID NO 130, 131, 132 or 133)

QHYRNVSGIYHV as for subtype 4l (SEQ ID NO 134)

IQVKNASGIYHL as for type 9 (SEQ ID NO 135)

AHYTNKSGLYHL as for subtype 7c (SEQ ID NO 136)

LNYANKSGLYHL as for subtype 7d (SEQ ID NO 137)

LEYRNASGLYMV as for type 10 (SEQ ID NO 138)

IYEMDGMIMHY or IYEMSGMILHA as for subtype 1d (SEQ ID NO 139 and 140)

VYEAKDIILHT as for subtype 1f (SEQ ID NO 141)

VWQLXDAVLHV as for subtype 2e (SEQ ID NO 142)

VWQLRDAVLHV as for subtype 2f (SEQ ID NO 143)

IWQMQGAVLHV as for subtype 2g (SEQ ID NO 144)

VWQLKDAVLHV as for subtype 2h (SEQ ID NO 145)

VWQLEEAVLHV as for subtype 2i (SEQ ID NO 146)

TWQLXXAVLHV as for subtype 2k (SEQ ID NO 147)

VYEADHHILHL or VYEADHHILAL or VFEADHHILHL as for subtupe 4k (SEQ ID NO 148, 149 and 150)

VYESDHHILHL as for subtype 4l (SEQ ID NO 151)

VFEAETMILHL as for type 9 (SEQ ID NO 152)

VYEAETLILHL as for subtype 7c (SEQ ID NO 153)

VYEANGMILHL as for subtype 7d (SEQ ID NO 154)

VYEAGDIILHL as for type 10 (SEQ ID NO 155)

VREDNHLRCWMAL or VRENNSSRCWMAL as for subtype 1d (SEQ ID NO 156 and 157)

IREGNISRCWVPL as for subtype 1f (SEQ ID NO 158)

ENSSGRFHCWIPI as for subtype 2e (SEQ ID NO 159)

ERSGNRTFCWTAV as for subtype 2f (SEQ ID NO 160)

ELQGNKSRCWIPV as for subtype 2g (SEQ ID NO 162)

ERHQNQSRCWIPV as for subtype 2h (SEQ ID NO 163)

EWKDNTSRCWIPV as for subtype 2i (SEQ ID NO 164)

EREGNSSRCWIPV as for subtype 2k (SEQ ID NO 165)

VREGNQSRCWVAL or VRTGNQSRCWVAL or VRVGNQSSCWVAL or VRVGNQSRCWVAL or VKEGNHSRCWVAL as for subtype 4k (SEQ ID NO 166, 167, 168 or 169)

VKTGNTSRCWVAL as for subtype 4l (SEQ ID NO 170)

IKAGNESRCWLPV as for type 9 (SEQ ID NO 171)

VKEGNQSRCWVQA as for subtype 7c (SEQ ID NO 172)

VKXXNLTKCWLSA as for subtype 7d (SEQ ID NO 173)

VRSGNTSRCWIPV as for type 10 (SEQ ID NO 174)

VKNASVPTAA or VKDANVPTAA as for subtype 1d (SEQ ID NO 175 and 176)

ARIANAPIDE as for subtype 1f (SEQ ID NO 177)

VSKPGALTKG as for subtype 2e (SEQ ID NO 178)

VSRPGALTRG as for subtype 2f (SEQ ID NO 179)

VNQPGALTRG as for subtype 2g (SEQ ID NO 180)

VSQPGALTRG as for subtype 2h (SEQ ID NO 181)

VSQPGALTKG as for subtype 2i (SEQ ID NO 182)

VSRPGALTEG as for subtype 2k (SEQ ID NO 183)

APYIGAPLES or APYTAAPLES as for subtype 4k (SEQ ID NO 184 and 185)

APILSAPLMS as for subtype 4l (SEQ ID NO 186)

VPNSSVPIHG as for type 9 (SEQ ID NO 187)

VPNASTPVTG as for subtype 7c (SEQ ID NO 188)

VQNASVSIRG as for subtype 7d (SEQ ID NO 189)

VKSPCAATAS as for type 10 (SEQ ID NO 190)

SPRMHHTTQE or SPRLYHTTQE as for subtype 1d (SEQ ID NO 191 and 192)

TSRRHWTVQD as for subtype 1f (SEQ ID NO 193)

APKRHYFVQE as for subtype 2e (SEQ ID NO 194)

SPQYHTFVQE as for subtype 2f (SEQ ID NO 195)

SPQHHNFSQD as for subtype 2g (SEQ ID NO 196)

SPQHHIFVQD as for subtype 2h (SEQ ID NO 197)

SPEHHHFVQD as for subtype 2k (SEQ ID NO 198)

RPRRHWTTQD or RPRRHWTAQD or QPRRHWTTQD or RPRRHWTTQE as for

subtype 4k (SEQ ID NO 199, 200, 201 or 202)
QPRRHWTVQD as for subtype 4l (SEQ ID NO 203)
RPKYHQVTQD as for type 9 (SEQ ID NO 204)
RPRMHQVVQE as for subtype 7c (SEQ ID NO 205)
RPRMYEIAQD as for subtype 7d (SEQ ID NO 206)
RHRQHWTVQD as for type 10 (SEQ ID NO 207)

or a part of said polynucleic acid which is unique to at least one of the HCV subtypes or types as defined in claims 2 to 3, and which contains at least one nucleotide differing from known HCV nucleotide sequences, or the complement thereof.

6. A polynucleic acid according to any of claims 1 to 5 having a sequence selected from any of SEQ ID NO 1 to 105, or a part of said polynucleic acid which is unique to at least one of the HCV subtypes or types as defined in claims 2 to 3, and which contains at least one nucleotide differing from known HCV nucleotide sequences, or the complement thereof.

7. A polynucleic acid according to any of claims 1 to 6, which codes for the 5' UR, the Core/E1, the NS4 or the NS5B region or a part thereof.

8. A polynucleic acid according to any of claims 1 to 7 which is a cDNA sequence.

9. An oligonucleotide primer comprising part of a polynucleic acid according to any of claims 1 to 8, with said primer being able to act as primer for specifically amplifying the nucleic acid of a certain isolate belonging to the genotype from which the primer is derived.

10. An oligonucleotide probe comprising part of a polynucleic acid according to any of claims 1 to 8, with said probe being able to act as a hybridization probe for specific detection and/or classification into types and/or subtypes of a HCV nucleic acid containing said nucleotide sequence, with said probe being possibly labelled or attached to a solid substrate.

11. A diagnostic kit for use in determing the genotype of HCV, said kit comprising a

primer according to claim 9.

12. A diagnostic kit for use in determining the genotype of HCV, said kit comprising a probe according to claim 10.

13. A diagnostic kit according to claim 12, wherein said probe(s) is(are) attached to a solid substrate.

14. A diagnostic kit according to claim 13, wherein a range of said probes are attached to specific locations on a solid substrate.

15. A diagnostic kit according to claim 14, wherein said solid support is a membrane strip and said probes are coupled to the membrane in the form of parallel lines.

16. A method for the detection of HCV nucleic acids present in a biological sample, comprising:

(i) possibly extracting sample nucleic acid,

(ii) amplifying the nucleic acid with at least one primer according to claim 9,

(iii) detecting the amplified nucleic acids.

17. A method for the detection of HCV nucleic acids present in a biological sample, comprising:

(i) possibly extracting sample nucleic acid,

(ii) possibly amplifying the nucleic acid with at least one primer according to claim 9, or with a universal HCV primer,

(iii) hybridizing the nucleic acids of the biological sample, possibly under denatured conditions, at appropriate conditions with one or more probes according to claim 10, with said probes being possibly attached to a solid substrate,

(iv) possibly washing at appropriate conditions,

(v) detecting the hybrids formed.

18. A method for detecting the presence of one or more HCV genotypes present in

a biological sample, comprising:

(i) possibly extracting sample nucleic acid,

(ii) specifically amplifying the nucleic acid with at least one primer according to claim 9,

(iii) detecting said amplified nucleic acids,

(iv) inferring the presence of one or more genotypes of HCV present from the observed pattern of amplified fragments.

19. A method for detecting the presence of one or more HCV genotypes present in a biological sample, comprising:

(i) possibly extracting sample nucleic acid,

(ii) possibly amplifying the nucleic acid with at least one primer according to claim 9 or with a universal HCV primer,

(iii) hybridizing the nucleic acids of the biological sample, possibly under denatured conditions, at appropriate conditions with one or more probes according to claim 10, with said probes being possibly attached to a solid substrate,

(iv) possibly washing at appropriate conditions,

(v) detecting the hybrids formed,

(vi) inferring the presence of one or more HCV genotypes present from the observed hybridization pattern.

20. A method according to claim 19, wherein said probes are further characterized as defined in any of claims 13 to 15.

21. A method according to claims 16 to 18, wherein said nucleic acids are labelled during or after amplification.

22. A polypeptide having an amino acid sequence encoded by a polynucleic acid according to any of claims 1 to 8, or a part thereof which is unique to at least one of the HCV subtypes or types as defined in claims 2 or 3, and which contains at least one amino acid differing from any of the known HCV types or subtypes amino acid sequences, or an analog thereof being substantially homologous and biologically

equivalent.

23. A polypeptide according to claim 22 comprising in its amino acid sequence at least one of the following amino acid residues:
I15, C38, V44, A49, Q43, P49, Q55, A58, S60 or D60, E68 or V68, H70, A71 or Q71 or N71, D72, H81, H101, D106, S110, L130, I134, E135, L140, S148, T150 or E150, Q153, F155, D157, G160, E165, I169, F181, L186, T190, T192 or I192 or H192, I193, A195, S196, R197 or N197 or K197, Q199 or D199 or H199 or N199, F200 or T200, A208, I213, M216 or S216, N217 or S217 or G217 or K217, T218, I219, A222, Y223, I230, W231 or L231, S232 or H232 or A232, Q233, E235 or L235, F236 or T236, F237, L240 or M240, A242, N244, N249, I250 or K250 or R250, A252 or C252, A254, I255 or V255, D256 or M256, E257, E260 or K260, R261, V268, S272 or R272, I285, G290 or F290, A291, A293 or L293 or W293, T294 or A294, S295 or H295, K296 or E296, Y297 or M297, I299 or Y299, I300, S301, P316, S2646, A2648, G2649, A2650, V2652, Q2653, H2656 or L2656, D2657, F2659, K2663 or Q2663, A2667 or V2667, D2677, L2681, M2686 or Q2686 or E2686, A2692 or K2692, H2697, I2707, L2708 or Y2708, A2709, A2719 or M2719, F2727, T2728 or D2728, E2729, F2730 or Y2730, I2741, I2745, V2746 or E2746 or L2746 or K2746, A2748, S2749 or P2749, R2750, E2751, D2752 or N2752 or S2752 or T2752 or V2752 or I2752 or Q2752, S2753 or D2753 or G2753, D2754, A2755, L2756 or Q2756, or R2757,
with said notation being composed of a letter representing the amino acid residue by its one-letter code, and a number representing the amino acid numbering as shown in Table 1,
or a part of said polypeptide which is unique to at least one of the HCV subtypes or types as defined in claims 2 to 3, and which contains at least one amino acid differing from known HCV types or subtypes amino acid sequences, or an analog thereof being substantially homologous and biologically equivalent to said polypeptide.

24. A polypeptide according to claim 22 comprising in its amino acid sequence at least one of the sequences represented by SEQ ID NO 107 to 207 as listed in claim 5, or part of said polypeptide which is unique to at least one of the HCV subtypes or types as defined in claims 2 to 3, and which contains at least one amino acid differing from

known HCV types or subtypes amino acid sequences, or an analog thereof being substantially homologous and biologically equivalent to said polypeptide.

25. A polypeptide having an amino acid sequence as represented in any of SEQ ID NO 1 to 106, or a part thereof which is unique to at least one of the HCV subtypes or types as defined in claims 2 to 3, and which contains at least one amino acid differing from known HCV types or subtypes amino acid sequences, or an analog thereof being substantially homologous and biologically equivalent to said polypeptide.

26. A recombinant polypeptide encoded by a polynucleic acid according to any of claims 1 to 8, or a part thereof which is unique to at least one of the HCV subtypes or types as defined in claims 2 or 3, and which contains at least one amino acid differing from known HCV types or subtypes amino acid sequences, or an analog thereof being substantially homologous and biologically equivalent to said polypeptide.

27. A method for production of a recombinant polypeptide of claim 26, comprising:

- transformation of an appropriate cellular host with a recombinant vector, in which a polynucleic acid or a part thereof according to any of claims 1 to 8 has been inserted under the control of the appropriate regulatory elements,
- culturing said transformed cellular host under conditions enabling the expression of said insert, and,
- harvesting said polypeptide.

28. A recombinant expression vector comprising a polynucleic acid or a part thereof according to any of claims 1 to 8 operably linked to prokaryotic, eukaryotic or viral transcription and translation control elements.

29. A host cell transformed with a recombinant vector according to claim 28.

30. A method for detecting antibodies to HCV present in a biological sample, comprising:

(i) contacting the biological sample to be analysed for the presence of HCV with a polypeptide according to any of claims 22 to 26,

(ii) detecting the immunological complex formed between said antibodies and said polypeptide.

31. A method for HCV typing, comprising:

(i) contacting the biological sample to be analysed for the presence of HCV with a polypeptide according to any of claims 22 to 26,

(ii) detecting the immunological complex formed between said antibodies and said polypeptide.

32. A diagnostic kit for use in detecting the presence of HCV, said kit comprising at least one polypeptide according to any of claims 22 to 26, with said polypeptide being possibly bound to a solid support.

33. A diagnostic kit for HCV typing, said kit comprising at least one polypeptide according to any of claims 22 to 26, with said polypeptide being possibly bound to a solid support.

34. A diagnostic kit according to claims 32 to 33, said kit comprising a range of polypeptides which are attached to specific locations on a solid substrate.

35. A diagnostic kit according to claims 32 to 34, wherein said solid support is a membrane strip and said polypeptides are coupled to the membrane in the form of parallel lines.

36. A pharmaceutical composition comprising at least one polypeptide according to any of claims 22 to 26 and a suitable excipient, diluent or carrier.

37. A method of preventing HCV infection, comprising administering the pharmaceutical compositon of claim 36 to a mammal in effective amount to stimulate the production of protective antibody or protective T-cell response.

38. Use of a composition according to claim 36 in a method for preventing HCV infection as defined in claim 37.

39. A vaccine for immunizing a mammal against HCV infection, comprising at least one polypeptide according to claims 22 to 26, in a pharmaceutically acceptable carrier.

40. A vaccine according to claim 39, comprising at least one polypeptide according to claims 22 to 26, with said polypeptide being unique for at least one of the HCV subtypes as defined in claims 2 or 3.

41. A peptide corresponding to an amino acid sequence encoded by at least one of the HCV polynucleic acids according to any of claims 1 to 8, with said peptide comprising an epitope being unique to at least one of the HCV subtypes or types as defined in claims 2 or 3, and with said peptide containing at least one amino acid differing from any of the known HCV types or subtypes amino acid sequences, or an analog thereof being substantially homologous and biologically equivalent.

42. A method for detecting antibodies to HCV present in a biological sample, comprising:

(i) contacting the biological sample to be analysed for the presence of HCV with a peptide according to claim 41,

(ii) detecting the immune complex formed between said antibodies and said peptide.

43. A method for HCV typing, comprising:

(i) contacting the biological sample to be analysed for the presence of HCV with a peptide according to claim 41,

(ii) detecting the immune complex formed between said antibodies and said peptide.

44. A. diagnostic kit for use in detecting the presence of HCV, said kit comprising at least one peptide according to claim 41, with said peptide being possibly bound to a solid support.

45. A diagnostic kit for HCV typing, said kit comprising at least one peptide according to any of claim 41, with said peptide being possibly bound to a solid support.

46. A diagnostic kit according to claims 44 or 45, wherein said peptides are selected from the following list:

- at least one NS4 peptide,
- at least one NS4 peptide and at least one Core peptide,
- at least one NS4 peptide and at least one Core peptide and at least one E1 peptide, or,
- at least one NS4 peptide and at least one E1 peptide.

47. A Diagnostic kit according to claims 44 to 46, said kit comprising a range of peptides which are attached to specific locations on a solid substrate.

48. A diagnostic kit according to claims 44 to 47, wherein said solid support is a membrane strip and said peptides are coupled to the membrane in the form of parallel lines.

49. A pharmaceutical composition comprising at least one peptide according to claim 41 and suitable excipient, diluent or carrier.

50. A method of preventing HCV infection, comprising administering the pharmaceutical composition of claim 49 to a mammal in effective amount to stimulate the production of protective antibody or protective T-cell response.

51. Use of a composition according to claim 49 in a method for preventing HCV infection as defined in claim 50.

52. A vaccine for immunizing a mammal against HCV infection, comprising at least one peptide according to claim 41, in a pharmaceutically acceptable carrier.

53. A vaccine according to claim 52, comprising at least one peptide according to claim 41, with said peptide being unique for at least one of the subtypes or types as defined in claims 2 or 3.

54. An antibody raised upon immunization with at least one polypeptide or peptide

according to any of claims 22 to 26 or 41, with said antibody being specifically reactive with any of said polypeptides or peptides, and with said antibody being preferably a monoclonal antibody.

55. A method for detecting HCV antigens present in a biological sample, comprising:

(i) contacting said biological sample with an antibody according to claim 54,

(ii) detecting the immune complexes formed between said HCV antigens and said antibody.

56. A method for HCV typing, comprising:

(i) contacting said biological sample with an antibody according to claim 54,

(ii) detecting the immune complexes formed between said HCV antigens and said antibody.

57. A diagnostic kit for use in detecting the presence of HCV, said kit comprising at least one antibody according to claim 54, with said antibody being possibly bound to a solid support.

58. A diagnostic kit for HCV typing, said kit comprising at least one antibody according to claim 54, with said antibody being possibly bound to a solid support.

59. A diagnostic kit according to claims 57 to 58, said kit comprising a range of antibodies which are attached to specific locations on a solid substrate.

60. A pharmaceutical composition comprising at least one antibody according to claim 54 and a suitable excipient, diluent or carrier.

61. A method of preventing or treating HCV infection, comprising administering the pharmaceutical composition of claim 62 to a mammal in effective amount.

62. Use of a composition according to claim 60 in a method for preventing or treating HCV infection as defined in claim 61.

Figure 1


```
                    1                                                  50
HCV-1          1a ATGAGCACGAATCCTAAACCTCAAAAAAAAAACAAACGTAACACCAACCG
HCV-J          1b --------A----------------G-----C-----------------
HCG9           1c ------------------------G-----C-----------------
BNL1           1d ------------------------G-----C-----------------
BNL2           1d ------------------------G-----C-----------------
CAM1078        1e ------------------------G-----C----A-A-----------
FR2            1f ------------------------G-----C------C-----------

HC-J6          2a --------A----------------G-----C----A-A-----------
HC-J8          2b --------A----------------G-----C----A-A-----A-----
S83            2c --------A----------------G-----C----A-A-----T-----
NE92           2d --------A----------------G-----C----A-A-----T-----
FR4            2f --------A----------------G-----CT---A-A-----T-----
BNL4           2e --------A----------------G-----C----A-A--T--------
BNL5           2h --------A----------------G-----C----A-A-----T-----

NZL1           3a --------ACT--------------G-----C----A-A-------T---
HCV-TR         3b --------ACT--------------G-C---C----A-A-----ACT---
NE48           3c --------ACT---A-----C----G-----C----A-A-------T---
NE274          3d --------ACT---A-----C----G-----C----A-A-------T---
NE145          3e --------ACT---A-----C----G-----C----A-A------GT---
NE125          3f --------ATT--------------G-C--CC----A-A-----ACC---

Z4             4a ------------------------G-----C-----------------
Z1             4b --------A----------------G-----C-----------------
GB358          4c ------------------------G-----C-----------------
DK13           4d ------------------------G-----C-----------------
GB809          4e ------------------T-----G-----C-----------------
BNL7           4k ------------------------G-----C-----------------

BE95           5a ------------------------G-----C----A-A-----------

HK2            6a --------ACT---A-----C----G-----C----A-A-----------

FR1            7a --------ACT---A-----C----G-----C----A-A--T--T-----

VN4            8a --------ACT---A-----C----G-----C----A-A-------T---
VN13           8b --------ACT--------------G-----C------A-----------

VN12           9a --------ACT---A-----C----G-----C----A-A-----A-----

NE98          10a --------ACT---------A----G-----C----A-A----------N
```

Figure 1 -continued


```
                51                                                  100
HCV-1       1a TCGCCCACAGGACGTCAAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAG
HCV-J       1b C--------------T-----------C--T------------------
HC-G9       1c C--------------T-----------------C-----------C----
BNL1        1d C-----T--K-GS--NNNNNNN-----------------------------
BNL2        1d C------------------N----------T------------------
CAM1078     1e C-------------------------C--T--C----------------
FR2         1f C--------------T--A-----------G--G--------G-------

HC-J6       2a ---------A-----T-----T-----C-----C-----------C----
HC-J8       2b C--------------T-----------------------------C----
S83         2c C-------------------------C--T--C-----------C----
NE92        2d C-------------------------C--T--C----------------
FR4         2f ---------------T-----------C-----C-----------C----
BNL3        2e C-------------------------C-----C-----------C----
BNL5        2h C--------------T-----------C--T--C-----------C----

NZL1        3a ---------------------------------A----------------
HCV-TR      3b ----------A----T-----------C-----A----------------
NE48        3c ---------------------------C----------------------
NE274       3d ---------------T-----------C-----C----------------
NE145       3e ------G--A-----T-----------C-----C----------------
NE125       3f C-------------------------C--T--G----------------

Z4          4a C-----CAT------A--------------T--C-----------C----
Z1          4b ------CAT---T--G--A--------C-----C-----------C----
GB358       4c C-----CAT------T-----------C--T--C-----------C----
DK13        4d C------AT------T-----------------C-----------C----
GB809       4e C-----CAT------T--------------T--C-----------C----
BNL7        4k C-----CAT------T--------------T--C-----------C----

BE95        5a -------------------------C--T--------------C----

HK2         6a -------AC-----------------------------------C----

FR1         7a ------TAT-----------------C-----C----------------

VN4         8a C-------------------------------C----------------
VN13        8b -------------------------------------------------

VN12        9a -------AT---T--------------C----------------------

NE98       10a C--G-----------T---------A--C----------------------
```

Figure 1 - continued

```
                 101                                                 150
HCV-1        1a TTTACTTGTTGCCGCGCAGGGGCCCTAGATTGGGTGTGCGCGCGACGAGA
HCV-J        1b -----C----------C--G----------T--G
HC-G9        1c ----------C---G----------G
BNL1         1d -----C----------C--GNN----------T--G
BNL2         1d -----C----------C--G----------C--G
CAM1078      1e -C---G--C-A----------AG--C-G
FR2          1f ----------C--G----------G

HC-J6        2a -A----------C--G----------A--G
HC-J8        2b --------C----------C--G----------A--G
S83          2c -A------C----------G----------G
NE92         2d -A----------CC-G----------G
FR4          2f ----------C--G----------C-A--G
BNL3         2e ----------C----------
BNL5         2h -A----------CC-G----------G

NZL1         3a -A---G----------AC----------C-T
HCV-TR       3b -A--TG--C-------T--------AC----------AGTAC-T
NE48         3c -A---G----------CT----------T--AC-T
NE274        3d -C-----AC----------A----------AGTTC-T
NE145        3e -A----------AC----------A--TC-T
NE125        3f -A---G-A----------AC----------AGT-C-T

Z4           4a ----------C--G----------TC--
Z1           4b --------C----------CC-G----------AG-TC-G
GB358        4c ----------C--G----------T--G
DK13         4d ----------T--G
GB809        4e ----------G----------TC-G
BNL7         4k ----------C--G----------TC-G

BE95         5a ----------GA----------TC-G

HK2          6a ----------CC-G----------

FR1          7a ----------C-T----------

VN4          8a -C------C----------GC-C----------
VN13         8b ----------C-T----------G

VN12         9a -C--------A----------AC-T----------G

NE98        10a -----G--C-A--A----------CCAG----------T--AGT-C-C
```

Figure 1 - continued

```
                151                                                   200
HCV-1       1a AAGACTTCCGAGCGGTCGCAACCTCGAGGTAGACGTCAGCCTATCCCCAA
HCV-J       1b ------------------------T--A--G--A--A-----------
HC-G9       1c ------------------------C--G--G----------T-----
BNL1        1d ----------------A--------T--C--G--A-------------
BNL2        1d ------------------G-----T-AC--G--A--------T--T--
CAM1078     1e --------G----------------T--G--G--C--A-----T-----
FR2         1f ------------------------C--A--G--A-------------

HC-J6       2a --------G--------C--G--A--T--A--G--C-----C-----T--
HC-J8       2b --------T-----A--C--G--G--T--AC----C-----C-----G--
S83         2c --A--------A-----C--G--A--T--G--G--C-----C-----T--
NE92        2d --A--------------C--G--A--T--G--G--C-----C--------
FR4         2f --------T--A-----C--G--A--T--A--G--C-----C-----A--
BNL3        2e --------T--A-----C--G--A--T--A--G--C-----C-----T--
BNL5        2h --A--------A-----C--G--A--T--G--G--C-----C-----T--

NZL1        3a --A-----T--A-----A--G-----C--AC----A--------------
HCV-TR      3b ------------------G-----CAAACAG-----C-T---------
NE48        3c ----------------A--G-----C-CGC-G--G--------------
NE274       3d --A-------------AG----C--CAACC-G--G--------------
NE145       3e --------A--------A-----C--C--AC-G--A--------T-----
NE125       3f --AT----------------------C--AC-G--G--------------

Z4          4a --------G----------------T--C--G-----A-----------
Z1          4b --------G--------A--------T--C--G----------------
GB358       4c --------G----------------T--G--------------------
DK13        4d --------G----------------T--G--G--C--------------
GB809       4e --------G----------------T--G--G--C--A-----------
BNL7        4k --------G----------------T--G-----C--A-----------

BE95        5a --------G--A-----------C--T--AC-G-----------T-----

HK2         6a --------------A--C--G--CA----C--G--C--A-----A--A--

FR1         7a -----C-----A-----C--G---A----C--G--C-----C--A--A--

VN4         8a --------T--A-----C--G--CA-------G--C--A--A--A-----
VN13        8b --A-----T--A-----C--G--CA-G--------C--A-----A--G--

VN12        9a --------G--A-----C--GG-CA-------G--C--A--A--A-----

NE98       10a ------------------------CA----G--C--A--C-------G
```

Figure 1 - continued 5/74

```
                 201                                                 250
HCV-1       1a GGCTCGTCGGCCCGAGGGCAGGACCTGGGCTCAGCCCGGGTACCCTTGGC
HCV-J       1b ------C----------T----------------------------------
HC-G9       1c ---C--C--A--------A---T-------G--------------------
BNL1        1d ------Y---Y----------T--------------------T--------
BNL2        1d ------C-A-T----T-----NN-------------A---C-T--C-----
CAM1078     1e --AG--C--A-----------T
FR2         1f ------C--A-----------T-------------T--------A-----

HC-J6       2a --A---G--CT--ACT----AAT------GAA-A--A--A-----C-----
HC-J8       2b A-A---G--CT--ACC----A-T------GAA----A--A--T--------
S83         2c A-A---G--CA--ACT----A-T------GAAG---A--A-----------
NE92        2d A-A---G--C---ACT----A-T------GAA-A--A--A-----------
FR4         2f A-A---G--CG--ACT----A-T------GA-GT--A--A-----------
BNL3        2e A-A---GN-NG--ACT------T------GA-GT--A--A--T--C-----
BNL5        2h A-A---G--CT--ACT----AAT------GA-GT--A--A-----------

NZL1        3a ---G------AG---A---C--T---------------------------
HCV-TR      3b ------CTC--G-------C--T---------------------------
NE48        3c ---G-----TGG------AC--T--------G------------------
NE274       3d ---A------AG-------C--T-------------T-------------
NE145       3e ---A--C-C-AG--GA--AC--T-------G-----T--------C----
NE125       3f ---A--C--AAG-------C--T-------C-----T-------------

Z4          4a ---G--C-A---A--------AT-------G--------------------
Z1          4b ---G--C---T----------T---------------------------
GB358       4c ---A-----AT-T-----A---T----------------------A----
DK13        4d ---G--C-AA-T------T---T-------------T-----T-------
GB809       4e ---G--C--AT----------AT-------G-----------T-------
BNL7        4k ---G-----AT-------A---T-------A-----A--A--T--A----

BE95        5a ---G--C-A----AC----C--T------G---A----------------

HK2         6a ---G--C-A----C--------CA---------------A----------

FR1         7a --TA--C-A---GACA---C-T-G-----G---A-----C----------

VN4         8a A-TG--C-AC-AAAC----C-T--------C--------------C----
VN13        8b --TG----AC-AAAC----C-T----------A----------C----

VN12        9a --TG--C-A-AA-C-A---C-A--------------T---*-----C----

NE98       10a ---G--C--AA----------T---------------------------
```

Figure 1- continued

```
                 251                                                  300
HCV-1       1a CCCTCTATGGCAATGAGGGCTGCGGGTGGGCGGGATGGCTCCTGTCTCCC
HCV-J       1b ------------C-----TATG--------A-------------A---
HC-G9       1c ------------C--------T----------------------C---
BNL1        1d ---------------------------------N----------C---
BNL2        1d -------------------------A--------------------C---
FR2         1f ---------CT--C-----------A--------------------C--T

HC-J6       2a ----A--C--G--------ACT---C-----A-------------C---
HC-J8       2b ----G--C--A--C-----T-----C--------T----------C---
S83         2c ----G-----G---------CT---C-----A--G----------C---
NE92        2d ----G--C--G---------CT---C-----A--G----------C---
FR4         2f ----G--C--G--C------CT---C-----A--G----------C---
BNL3        2e ----G-----G--C-----GCT---C-----A-------------C---
BNL5        2h ----G-----G--C------CTT--T-----A---------T----C--T

NZL1        3a ----------T--C-----------------A--G----------C--A
HCV-TR      3b -------C--G---A-------T----T---A---------T----C---
NE48        3c -------C--T-------------------------------------C---
NE274       3d -T--T-----------------T--------A---------T----C---
NE145       3e ----------T--C-----------------A--G-----T--------T
NE125       3f ----------G----------T--------A-----------------

Z4          4a -------------------------------A--G-------------T
Z1          4b ----T--C--------------T--------A--G----------C---
GB358       4c -T--T--C--T--------T-------------------------A--T
DK13        4d ----T--C-------------------------------------A---
GB809       4e ----T--C--------------T--------A--G----------C--T
BNL7        4k -T--T--C--T--------T----------ANN-------T----C---

BE95        5a ----T--C-C----------CT---------A--G-----G--C--C--T

HK2         6a -T--T-----A--C--------T--------A--T----------C---

FR1         7a ----T--------C----------A--------------------C---

VN4         8a -T--T-----A----------T--T-----A--C----------C---
VN13        8b -T--T-----G---------T--T--C-----A--G----------C---

VN12        9a ----T-----G--C----------C--------G------T----C---

NE98       10a ----A-----G--------------------A--G----------C--G
```

Figure 1 - continued

```
                 301                                                  350
HCV-1       1a CGTGGCTCTCGGCCTAGCTGGGGCCCCACAGACCCCCGGCGTAGGTCGCG
HCV-J       1b ----------------T------------------------------
HC-G9       1c --C-------------T--------TT-T-----------G-----A--
BNL1        1d --C-------
BNL2        1d --C-------
FR2         1f --C-----C--T----------------AT--------A-----A--A--

HC-J6       2a --A--T--C--T--CTCT----------AT----------A------C--
HC-J8       2b --C--G-----T----CT----------C----------A---A--A--
S83         2c --C--T-----C---TCA----------C----------A--AA-----
NE92        2d --A--G-----C--GTCA--------A--T----------AC-----A--
FR4         2f --G--------C--CTCG--------A-AC----------AC-----A--
BNL3        2e --A-------
BNL5        2h --A-------

NZL1        3a --C-----C--T--ATC---------A-AT----------G-----C--
HCV-TR      3b -----T-----C-----T--------A-AT--------------A--C--
NE48        3c --C--T--------G----------A-AT----------A--A--C--
NE274       3d --C----------ATCT----------AT----------A-----T--
NE145       3e --C-----C--A--G--T----------AC----------A-----C--
NE125       3f --------C--C-----T--------A-AT--------------A--A--

Z4          4a --C----------ATCT--------A-AT--T--------G--A-----
Z1          4b --C--T--CA----GTCT----------AT--T--------------C--
GB358       4c ----------A--GTCT--------A-AT--T--------A-----C--
DK13        4d --------------GTCT--------G-AT--T--------G-----C--
GB809       4e --C--G--------GTCT--------T-AT--T--------G-----C--
BNL7        4k --C--T----

BE95        5a --A-------------AT----------AT----------A-AA-----

HK2         6a --C-----C-----ACAT----------AT----------C-A--C--

FR1         7a --C--G-----T----AT----------AC----------A-----C--

VN4         8a --C--------C--A-AT--------A-AC----------G-----C--
VN13        8b -NC--------C----AT--------T-AT---------N-G-----C--

VN12        9a ----------C--GGA------N----AT---------N-G-----C--

NE98       10a --C-------
```

Figure 1 - continued

```
                 351                                                  400
HCV-1       1a CAATTTGGGTAAGGTCATCGATACCCTTACGTGCGGCTTCGCCGACCTCA
HCV-J       1b T----------------------------A------------------
HC-G9       1c ----------------------------C-----T---------------
FR2         1f ----------------------------A-----T----------T----

HC-J6       2a ---CG----------------------A----------T----------
HC-J8       2b ---------C-GA-------------A-------T--T--T----------
S83         2c ---C-----------------------A--------T--T----------
NE92        2d ---C-------------------------------T-----T----------
FR4         2f ---C-----------------------C-----T-----T-S--------
BNL3        2e                                 --------N-NT----------

NZL1        3a ------------A--------------A--------A-------------
HCV-TR      3b ---C--T--------------------A-----T--A-------------
NE48        3c ------------------------------A-----G-------------
NE274       3d ---CC-------A--------------A--------A--------T----
NE145       3e --------------------C--T--C--A-----G--------T----
NE125       3f ---C-----------------------C-----T--A--------T----

Z4          4a ----C----------------------G----------------------
Z1          4b T---C-------A--------------G-----T----------------
GB358       4c ---C-----------------------A--C--------T----------
DK13        4d ---C-----------------------A--T------------------
GB809       4e ---CC----------------------A--A------------------

BE95        5a T--------------------------A--------A--------T----

HK2         6a G--------------------------A-----T--G--------T----

FR1         7a ---C--------------------------A-N---NC-A----------

VN4         8a ---C--------A--------C--------T------------------
VN13        8b ---CC-------------------------T--N--S-------------

VN12        9a ---CC----------------C-----C--T------------------
```

Figure 1 - continued

```
                401                                                      450
HCV-1       1a TGGGGTACATACCGCTCGTCGGCGCCCCTCTTGGAGGCGCTGCCAGGGCC
HCV-J       1b ----------T-----T----------C--A--G---------------
HC-G9       1c ----------C--------------T-----A--G-----------A--T
FR2         1f ----------T----------------C--A--G------T----AA--

HC-J6       2a ----------C--TG----A--------G--C--C----TC-----A--T
HC-J8       2b ----------C--TG----T--------GG---------TC-----A--T
S83         2c ----------CG----T-----T--CG----C----T------A---
NE92        2d ----------C--TG-------------AG----T--T-TC-----A--T
FR4         2f ----------TG--------------G-G--C----T------A---
BNL3        2e ----------N--CG-T----------GG-G--C--G-TN---------

NZL1        3a ----------C--------------T---G-A-------TC--A--A---
HCV-TR      3b ----------T----------------G-G--G----TC--A--A---
NE48        3c ----------T----------------CG-G--G----T---A------
NE274       3d ----------T--------------T---G-A--G----TC--A--A--T
NE145       3e -------T--T--------------T--GG-A-------TC--G------
NE125       3f ----------T--------T--T-----CG-A--G----TC--A------

Z4          4a ----A-----C---A----G--------CG-G--G----TC--------T
Z1          4b ----A-----T--------A---------G-G--T----TC---------
GB358       4c ----A-----C--------A--------CG-G--T----TC---------
DK13        4d ----A-----C---G----A--------CG-G--T----TC-----A---
GB809       4e ----A-----C-----T--A--------CG-G--T----TC-----A---

BE95        5a -------T--C--------A----G---CA----G----TC--A-----T

HK2         6a ----------T--CG----G-----G---T-G--C----TC--GGCT--G

FR1         7a ----------C--TG--C-A--A-GG--G-----C----T---GGCT---

VN4         8a -------T--C--TG----A-----T--GW-G-------TC--GGN----
VN13        8b -A-A------T--

VN12        9a ---A------C--TG----T--------C----------T---GGC--AA
```

Figure 1 - continued

```
                451                                                 500
HCV-1       1a CTGGCGCATGGCGTCCGGGTTCTGGAAGACGGCGTGAACTATGCAACAGG
HCV-J       1b -----A-----T-------------G----------------------
HC-G9       1c -----A-----T--TA-A--C-----------------T--C--------
BNL1        1d                          ----------------------
BNL2        1d                          ----------------------
FR2         1f ---N-A-----T--------C----N--G--------TNNNNNNNNNNNN

HC-J6       2a --C-----------GA-A--C-----G-----G--T--T-T---------
HC-J8       2b -----A--C--T--TA----C-----G-----GA-A--T--C--------
S83         2c --C--C-----G--GA----------G-----GA-A--T--------G--
NE92        2d --C-----------GA-A--------------GA-A--------------
BNL3        2e --C---N-------G-----C-----G-----GA-A--T----N------
FR4         2f --C-----------G-----C-----G-----GA-A--T-----------
BNL4        2g                          -----G--A--T-----------
BNL5        2h                          -----GA-A-----C--------
BNL6        2i                          -----GA-A--------------

NZL1        3a --C-----------GA---CC--T--------GA-A--T-TC--------
HCV-TR      3b --C--T-----T--GA---CA--T-GG-----A-----------------
NE48        3c --C-----------GA---C---T--G-----GA-T----TC--------
NE274       3d --C--A-----T--GA-A-CC--T--G-----AA-A--T-TC--------
NE145       3e --C--A--C--G--AA---C---C--G-----AA-A--T-T---------
NE125       3f --A--A-----T--GA---C---T--G-----AA-A----T---------

Z4          4a ---------------A---C-G----G-----GA-T--------------
Z1          4b ---------------A---CCG----G-----AA-T-----C--------
GB358       4c -----A--C--T--TA---C-G----G-----GA-C--T-----G-----
DK13        4d ---------------A--C-------G-----G--C--T-----------
GB809       4e -----A--C--T--TA---C-G----------GA-C-----C--------
BNL7        4k                          -----GA-C--T-T---------
BNL8        4k                          -----GA-C--T-----------
BNL9        4k                          -----GA-T--T-----------
BNL10       4k                          -----GA-C--T-----------
BNL11       4k                          -----GA-T--T-----------
BNL12       4l                          -----GA-C--T-----------

BE95        5a --C--A--C--T--GA----C--T--G-----G--A--------------

HK2         6a --C--A--------GA---CAA-C--G-----GA-C--T-----------

FR1         7a --------------TA---CAA-C--G-----G--C--T--C--------

VN4         8a T----------G---AN--NCA-C--G-----N--A--T--C-------N

VN12        9a ----NA-----T---A---CCA-C--G-----GA-A--------------

NE98       10a                          -----AA-T--T-TC--------
```

Figure 1 -continued

```
                 501                                                  550
HCV-1        1a GAACCTTCCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCCCTGCTCTCTT
HCV-J        1b ---T--G--C----------------------CT-A--TT----G----
HC-G9        1c ------C--C----------T---------T-G--C--T--T--A--C-
BNL1         1d ----T-G--C----------------------CT----TT----G--C-
BNL2         1d ---TT-G-------------------------CT-A--TT-T--G--C-
FR2          1f N--------N----------NN----------CT----NT-A-------

HC-J6        2a ----T-A--C--------C--T---------T-G----------G--C-
HC-J8        2b ---TT-A--C----------T--------TT-G--T--T--T--G--A-
S83          2c ---TT-G--C--------------------T--CT-------CT-G----
NE92         2d ----T-G--C--------C--T---------T-AT----------A----
BNL3         2e ------C----------C--T---------TNGT----T--T--G----
FR4          2f ----T-G--C--------C--T---------T-G-----T--CT-G----
BNL4         2g ---T--G----------------------T-GT----T--T--G----
BNL5         2h ---T--G--C--------C--T---------T-G------T----A--C-
BNL6         2i ------G----------C--T---------T-A----------T----

NZL1         3a ----T-G--C--------C--T--------------T--T---T------
HCV-TR       3b ---T--------------C--T-----T-----C--C--T--CT----C-
NE48         3c ---TT-A----------C--T---------T-G--T--T--CT----A-
NE274        3d ---TT-A--C--------------------T-G--T--TT---------
NE145        3e ---------C----------T---------T-G--T--T-----G--A-
NE125        3f ---TT-G--C--------C--T--------------T--T--CT----A-

Z4           4a ---T-----C----------------------T----A--T--T--G-
Z1           4b ----------------------------T-----T--A--T-----G-
GB358        4c ---T-----C--------------------T-CT----A--T--T--G-
DK13         4d ---T-----C----------------------CT----A--------G-
GB809        4e ---T--C--C--------C--T----------CT----A--T-----G-
BNL7         4k ------C--C--------C--T----------CT----A--C-----G-
BNL8         4k ---------C----------T----------CT----A--C-----G-
BNL9         4k ---T-----C--------C--T----------CT----A--T-----G-
BNL10        4k ---TA----C--------Y--T----------Y----A--T-----G-
BNL11        4k ---Y--C--C----------T----------CT----A--T-----G-
BNL12        4l ------C--C--------------------A-C-----A--T-----G-

BE95         5a ---TT-A--C--------------------TA----T--T--T-----G-

HK2          6a ---T--C--C----------------------T----A--A-----G-

FR1          7a ---T--------------C--T----------CT-A--A---T-A--G-

VN4          8a ---T-----C--NN-----N----------N--CT----A--T-----G-

VN12         9a ---T----------------------------WCT----A--T-----G-

NE98        10a ---TT-A--------------------------------TT--T----A-
```

Figure 1 - continued

```
                 551                                                  600
HCV-1        1a GCTTGACTGTGCCCGCTTCGGCCTACCAAGTGCGCAACTCCACGGGGCTT
HCV-J        1b -T-----CA-C--A-----C--T---G-G---------GTGT-C---A-A
HC-G9        1c --C----A--C--T---------GT-GG-----------TT-----G-G
BNL1         1d -------G--T--AA-KA-C--TC--G-G---------G-AT-C---G-G
BNL2         1d -------G--T--AA--A-C--TC-TG-G---------G-AT-C---G-A
FR2          1f --C-C--A--C---A-C--T-----TG-G----A--G-A-A--C-ATGGC

HC-J6        2a --A-C--CACC--G-TC--C--TGC-G-----AAG---AT--GTACCGGC
HC-J8        2b --G-C--A-----A-TG--T--AGTGG----CA-G---ATT-GTTCTAGC
S83          2c --A-CT-------A-T---C---GTGG-G--CAAGG--A--GGC-ACTCC
NE92         2d -TA-C--------G-TC--C-G--TG--G--CAAG---A---GCA-CTC-
BNL3         2e -TG-C--C-----T-TC--T-N-GTTG-G--CAAA--TA---GTCA-GCC
FR4          2f -TA-C--C-------TG--T---ATA--G--TAAG---AA--GCCACT-C
BNL4         2g -TG-C--C-----T-TC--T---GTG--G--TAAG---A---GTACCA-G
BNL5         2h -TC-C--------G--G--C--TGTG--G--CAAG---A---GCCACTC-
BNL6         2i --A-C--C-----G-TC--T---GTG-----TGCG---CG--GT--TTC-

NZL1         3a ----A-T-CAT--A--AG-CAGTCTAG-GTG---G--TA-GT-T--C--C
HCV-TR       3b --------TGC-----G--T-G--TAG-GTACACG---A-GT-T--C--A
NE48         3c -----GTCTGT--T--AG-A-GGCT-G-GTAC--G--TGTAT-C--C--C
NE274        3d -----GTCTGT--T---G-A-GGATTG--TAC--G--TGTGT-T--C--C
NE145        3e -----CT-TGC--T--AGTC-GG-TGG-G--T------G-AT-C--T--C
NE125        3f -----GT-TCC-----AG---GGCTAG-GTACA-G---A-GT-C--C--A

Z4           4a --C-C-----T--A--G-----TG-G--CTAC--G--TG-TT----CA-C
Z1           4b --C----AACA--A--A--T---GTG--CTAC--G--TG-TT----CG-C
GB358        4c --C-------T---A-C------GT-A-CTAT-----TG--T----CA-C
DK13         4d --C-------T---------------A-CTAT------AG-T----TG-C
GB809        4e --C-C-----T-----G----G-GTTA-CTAT-----TG-TT----CG--
BNL7         4k --C-------C-----------AT-A-CTAT-----TGT-T----CA--
BNL8         4k --C-------T-----------ATTA-CTAC------A--T----CA-C
BNL9         4k --C-------C-----------ATTA-CTAC-A----A--T----CA-C
BNL10        4k -TC-------C-----------ACTA-CTAT------GT-T----CA-C
BNL11        4k --C-------C-----------AC-A-CTAC-----TGT-T----CA--
BNL12        4l --C-------C--G--C-----TC-G--TTAT--G--TGT-T----CA--

BE95         5a -TC----C--T--G--C--T--AGTT-CCTAC--A--TG--T-T---A--

HK2          6a --C-C--AAC---A--------TCTTACCTACG---------GT-----A

FR1          7a --C-C---ACA--A--C--A--AATT-----CAAG---G--T-T---A-C

VN4          8a --C-T--AACA--A--C--C--GGCG--TTATAC----AAGT-T--C--G

VN12         9a --C-C--CAC---T--C--C--ACTAA-CTATGCT---AAGT-T-----G

NE98        10a -----CT-ACA---A-AG-C-GGCTGG-GTAC--T--TG--T-C--A--C
```

Figure 1 - continued

```
                 601                                                  650
HCV-1       1a TACCACGTCACCAATGATTGCCCTAACTCGAGTATTGTGTACGAGGCGGC
HCV-J       1b -----T-----G--C--C---T-C-----A----------T-----A--
HC-G9       1c -----T----------------C--TG--TCCG----------A---A
BNL1        1d --T--T--------C--C--TT-C--------C--CA-C--T---AT--A
BNL2        1d --T--TC----------C--TT-C--------C--CA-C--T---AT-AG
FR2         1f -----T-----T-----C--TT-C---GGC--C--C--A--T-----AAA

HC-J6       2a ---ATG--G-----C--C---A-C--TGAT--C---ACC-GGC-ACTCCA
HC-J8       2b ---T---C---T---------T-A---AAC--C--CACC-GGC--CTCA-
S83         2c ---ATGCCG-----C------T-C-----T-----C--T-GGC--CTT-A
NE92        2d ---ATG--A--------C----AG---AGT--C--C--C-GGC--CTCAG
BNL3        2e --TATG-CA-----C--C---T-C---AAC--C--C--A-GGC-ATT--N
FR4         2f ---ATG-CG--T-----C--TG-C--TGAC--C--C--C-GGC--CTCAG
BNL4        2g ---ATG-CA--------C--TT-C---AAC--C--CA-C-GGC-AAT-CA
BNL5        2h --TATG--G----------T-A---AGC--C-----C-GGC--CTTAA
BNL6        2i ---ATG--G----------T-G---AGC--C--C--T-GGC--CTC-A

NZL1        3a ---GT-C-T-----C--C--TT-C--TAGC----------T-----C-A
HCV-TR      3b --TGTGC-T-----C--C---T----TGG---C-------------C-A
NE48        3c ---ATAC----------C--TT-G---AGC--C--A-----T-----C-A
NE274       3d ---GTGC-------C--C---T-----GGC-----C-----T-----CC-
NE145       3e ---ATGC----------C---T-A---AGC--C--A--A--T-------A
NE125       3f ---ATAC-T-----C--C---T-----AGC--C--C-----T-----T-A

Z4          4a --T---A----------T--G--T--C-----A--C--T--A--T-A
Z1          4b --T--T--------------------A-C--C--A---------A---A
GB358       4c --T---A----------C-----G--------C--A---------A-C-A
DK13        4d -----T--------C--------G--------C--A--C--T--AA-C-A
GB809       4e --T---A-------C--C-----G--TG----C--A---------A-C-A
BNL7        4k ---T-T----------------G--T--A--C--A-----T-----C-A
BNL8        4k ----------------C-----G--------C--A--T--T-----C-A
BNL9        4k --T--TA-------C--C-----G--T--A--C--A-----T-----C-A
BNL10       4k -----T----------C-----G--T--A--C--A-----T-----C-A
BNL11       4k -----T----------C-----G--T--A--C--A----TT-----C-A
BNL12       4l ----------C--C-----G-----C--C--A-----T---T-C-A

BE95        5a --T--T--T----------A-----TTCC--A--C--T-----A-A

HK2         6a -----TC----A----------C-----C--C--C---CTG-------A

FR1         7a -----TC-T--------C---T-G---AAC--C--C--T-TT-------A

VN4         8a -----TC-------C--C-----C---AGC--C--C--T--T-------A

VN12        9a --T--TC-A--------C-----C--TAGC--C--------T------AA

NE98       10a ---ATG--A--T--C--C---AG----GGT-----C-----T-----C-G
```

Figure 1 - continued 14/74

```
                 651                                                  700
HCV-1       1a CGATGCCATCCTGCACACTCCGGGGTGCGTCCCTTGCGTTCGTGAGGGCA
HCV-J       1b G--CATG---A-------C--C--------G--C-----C--G---A-T-
HC-G9       1c GA-CCTG---A----TCTG--C-----T--G--C-A---A--C-------
BNL1        1d --G-ATG---A-----TAC--A--------G--C--------G----AT-
BNL2        1d T-G-ATG---T-----G-C--A-----T--G--C--------G---AA--
FR2         1f G--CAT------T-----G--T-----N--G--C---A-A--G--A----

HC-J6       2a G-C---TG----C---GTC--C----------G----AGAAA-T---G-
HC-J8       2b T--C--AG-T--C--TCT---T--A--------A--T-AGAA---TAATG
S83         2c A-GA--AG-G--T--T-----T--A----------T-AG---ACC-C--
NE92        2d G-----TG-T--T---GTC--C-----T--------T-AGGAGA------
BNL3        2e G--C--GG-G--T--TGT---T--A--T-----C----AGAA-AGCTC-G
FR4         2f G--C--GG-G--C--TGT---T--A--T-----C--T-AGA-GTCA--T-
BNL4        2g G-GC--GG-G--T--TGT---T--A--T-----G--T-AGTTGC------
BNL5        2h G-----TG-G--T---GTC--T--A--T--T--A--T-AGA-GC-CCAA-
BNL6        2i G--G---G----T---GTC--T--A--T--T--C--T-AGT-GA---A--

NZL1        3a T----T---T--------A--C--C--T--A-----T--C-AG--C----
HCV-TR      3b A----TG---T-----TTA--C--A-----G--C-----CACAACC----
NE48        3c -C---T----T-----TTG--C--T-----A--C-----C-AAA-CAAT-
NE274       3d T--A-T----T-----TTG--A--T--T--G--C-------AATCA----
NE145       3e A----TG----------TG--T--T-----T--C-----G-AGA-C----
NE125       3f TA---T----------TG--C--C--T--G--C---AC---C-----T-

Z4          4a -C--CA------A---TTG----------A--C--T--GATGACT--G-
Z1          4b GC-CCA----A-----TTG--A-----T-----C--T--G--GAC--AG-
GB358       4c GC-CCA------A---CTC--A-----TT-A--C-----GA-G-TT--G-
DK13        4d TT-CCA----T-A---CTC-----A-----T--------GA-G--A--G-
GB809       4e -A--CA----T-A---CTC--A--------A--C--T--GAAGACC--G-
BNL7        4k -C--CA----T-----CTC--A--T-----G--C-----GA-A-----G-
BNL8        4k -C-CCA----T-----CT---A--T-----G--C-----GA-AACT--G-
BNL9        4k -C--CA----T----TCTC--A--T-----G--C-----GA-A-T---G-
BNL10       4k -C--CA----T-AGCACT---A--T-----G--C-----GA-A-T---G-
BNL11       4k -C--CA----T-----CT---A--A-----G--C-----GAAA-----A-
BNL12       4l -C--CA----T-A---CTA--A-----T--A--C--T--GAAGACT--G-

BE95        5a TA-CCTG-----A---G-A--T--T-----G-----T--CATGACA--T-

HK2         6a T-C-ATG---T----TTTG--T--A---T-G-----T--GA-G-TC-ATG

FR1         7a GACCATG--A-----TCT---A--T--T-----A--TA-CAAG-C---G-

VN4         8a GACACTG--TT-----TTG--T-----T--A-----T--GAAGRT-RA--

VN12        9a T-GCATG--------TCTC-----T--------C-----GAAGACC----

NE98       10a G---ATT-----C---TTA--T--C--T-----C-----A--CTCT----
```

Figure 1 continued

```
                 701                                                  750
HCV-1       1a ACGCCTCGAGGTGTTGGGTGGCGATGACCCCTACGGTGGCCACCAGGGAT
HCV-J       1b -TTT---CC-T--C-----A---C-C--T--C---C-C--GG-----A-C
HC-G9       1c ------------------CT-CC-T-GT--C--C--A---G--------
BNL1        1d --CATCTCC-C--C---A-----C-C-----C---C-T--GGT--AAA-Y
BNL2        1d --T-T--TC-T--C---A--C-RC-C-----C---C-T--GGT--AA--C
FR2         1f -TAT---CC-T--C-----AC--C-C-----C---C-C--AG-GC--ATC

HC-J6       2a -TA-A--TC----C---A-AC--G-CT-A--G-AT------GTGCA-C-G
HC-J8       2b G-A---T-CAT--C---A-ACAAG-A--A--C-AC-----TGTG-AAC-C
S83         2c ---T---TC-A---------C--G-TG----C-ATC-C----TA--TC-A
NE92        2d --ATA--CC-C------A-AC--G-TT-G--C-ATA-A--TGTG--CC-A
BNL3        2e GTCGG-TCCAC------A-CC----CT-G--C-ACA-A---GTG--CA-A
FR4         2f -TAGGA-CTTC------ACA---G-CT-G--C-AC-----TGTG--CCGA
BNL4        2g -TAAG--CC----C---A-AC--G-C--T--C-AC-----TGTG-ACC-G
BNL5        2h -TCAG--TC-C--C---A-AC-TG----A--C-AT------GTG--CC-A
BNL6        2i --A----CC-C--C---A-AC--G-C-------ACA-C--TGTG--CC-A

NZL1        3a -TA-A--T-C---C---ACCC-AG----A-----A-----AGT----T-C
HCV-TR      3b --CAA--ATCA--C---ACAA--G-CT-AA-G---------GTT---ACC
NE48        3c --A--A---C---C---A-AC--G----T--G--A-----GGT---TC-C
NE274       3d --T-----CAA--C---A-TC--G--G-A--A--A-----GGTT-A-T-C
NE145       3e --A-A---GA---C---ACCC--GC---A--A--------AGT---AT-C
NE125       3f --CAG--A-----C---AC-C-AG-A--A--G--A-----TGT--AAC--

Z4          4a --A-A---C-T--C---AC-C--G----G-----A-----TGT-GCAC-C
Z1          4b -TA-T--TC-C--C------C-CT-------C--T------G-GCCCT--
GB358       4c -TCAG--AC-C--C--------CC-C--T--C--C-----GG-GCCTT-C
DK13        4d --AAG--T-CA--C------T-TC-C-----C--C-----TG-GCAAC--
GB809       4e --CAG---C-------------CC-C--T--C--A-----GT-GCCTT-C
BNL7        4k -TCAG--AC-T--C-----A--CC-T--------C--C--AG-GCCAT-C
BNL8        4k -TCAG--AC-T--C--------CC-T--T-----C--C--AG-GCCAT-C
BNL9        4k -TCAG-----T--C--------CC-T--------CA-C--AG-GCCAT-C
BNL10       4k --CAG--AC-C--C--------CC-T--------C--C--AG-GCCAT-C
BNL11       4k -TCAT--AC-C--C--------CC-T--------C--C--AG-GCCAT-C
BNL12       4l --A-T---C-C--C--------CT-A--A-----C------G-GCCCATA

BE95        5a -T-TGAGT--A--C-----CCAA--T--------AC--T-AG--CC-AGC

HK2         6a -TCGG--C-CC------CAT--TG-------C--CC------TACCAA--

FR1         7a -T-AG--AC-A------C-CC-TG-CT----C--CT-A---GT-CCCA-C

VN4         8a -TCAA--CC----C------CA-GCCT----G--CC----AGTGCC-A-C

VN12        9a --CTGA-C-A-------C--T--GCCT----G--AT----GGTGCA-A--

NE98       10a -TA-A--A--A--C---A-CC-TG---G---Y--C--C---GTG-A-TCG
```

Figure 1 - continued

```
                751                                                    800
HCV-1       1a GGCAAACTCCCCGCGACGCAGCTTCGACGTCACATCGATCTGCTTGTCGG
HCV-J       1b A---GCA-----A-C---ACAA-A-----C---G-----T----C--T--
HC-G9       1c TCGCGCG------TC-GTG--G----G----GTG----CTC-A-------
BNL1        1d -CT-GTG-----A-TR--GCAA-C---------G----CT-------T--
BNL2        1d -CT--TG----TA-TG--GCAA-C-----C--TG----CT----G--T--
FR2         1f -CG--CGCT---ATCGATG--G-G--G------G----C--C--C--G--

HC-J6       2a CC-GGCGC--T-A--CA-GGCT-A--GACG-----T--CA--G----GAT
HC-J8       2b C--GGTGCG-T-A-TCGTAGC--G---ACA---G----CA--A-C--AAT
S83         2c CCTGGCGCT-T-A-T-A-GGC--G---GCA---------A-CA-C--GAT
NE92        2d CCTGGTGCG-TTA-C-A-GGC--G--GACG--T--T---ACCA-CA-T-C
BNL3        2e CCTGGTGCT-T-A-C-A-GGA--G--GGCA-G---T---GCCG-C--GAT
FR4         2f CCTGGTGCT-T-A-T-GAGGT--G--GGC------T---ACCA-C--GAT
BNL4        2g CC-GGCGC--T-A-T-G-GGCT-G--GACG-----T--CACCA-C--GAT
BNL5        2h CCTGGCGCG-T-A-C-G-GGTT-G--GACG-----T--CACCA-C--T-C
BNL6        2i CCTGGCGCG-TTA-C-A-GGC--G--GACA--T--T--CA-CA------C

NZL1        3a -T-GG-GCAA-TA-TG-TTC-A-A--CA----TG-G--C--AT-A--A--
HCV-TR      3b CTTGGCG-GA--A-CG--TC-A-C---ACC--TG-G---A----G--A--
NE48        3c -T-GGTGCGA--A-CG-ATC-A-C--CG-G---G-G--------G--G--
NE274       3d -CTGGCGCGA--A-TG-ATC-A-C--CA----TG-G--------G--G--
NE145       3e -CTGGTGCAA-GA--G-TTCCG-A--CG-A---G-G---T----A-----
NE125       3f CCTGGCGCAGT-A-CG-ATCAA-C--CA-G--TG-G---T--A-G--G--

Z4          4a CCGGGCGCT--GCTTGA-TC-T-C--G--A--TG-G--CT-AA-G--A--
Z1          4b CC---CGCA--GTTAGA-TCCA-G--CA-G--TG-A--C---A-G--G--
GB358       4c AT-GGCGCT--GCTTGAATCC--C--GA----TG-G------A-G--A--
DK13        4d CTG--TGCT--GCTTGA-TCTT-GA--------G-G------A-G--G--
GB809       4e -T-GGTGCT--GCTCGA--CCT-G--G--C--TG-G--C---A-G--A--
BNL7        4k AT-GGCGCG--ACTTGA-TCT--A--GA----TG-G--CT--A-G--G--
BNL8        4k AT-GGCGCA--GCTTGA-TCT--G--GA----TG-G------A-G--G--
BNL9        4k AT-GGCGCA--GCTTGA-TCCT-G--GA----TG-G------A-G--G--
BNL10       4k AC-GCGGCG--GCTTGA-TCC--G--GA----TG-G------A-G--G--
BNL11       4k AT-GGCGCG--ACTTGA-TCT--A--GA----TG-G---G--A-G--G--
BNL12       4l CTTTCGGCT--ACTT-T-TCCG-A--G--G--TG-G------A-G--G--

BE95        5a CT-GG-GCAGT-A--G-T-CT-----GA-AGC-G-T--CTAC--A-CG--

HK2         6a -CTTCCACG-----A---GGAT-C--CA-G--TG-G-----T----CG--

FR1         7a TCATC-G-G--AATCCACGG-T----C--A---G-A--C--C--C--T--

VN4         8a -CGTCTACG--A-TC--CGG-T-C--CAAA--TG-G--CA-CA-G--G--

VN12        9a -CGTCGG-GT--ATC-G-GGTG-C--CGAG---G-G--C--CT-G--G--

NE98       10a CC-TGCGC-G--A-CG-CTCT--C--CACG---G-G---A--A-G--G--
```

Figure 1 - continued

```
                801                                                      850
HCV-1       1a GAGCGCCACCCTCTGTTCGGCCCTCTACGTGGGGGACCTATGCGGGTCTG
HCV-J       1b -GCG--TG-T--------C--TA-G-----T-----T--C-----A--C-
HC-G9       1c -GC---TG-GT----------TA-G--T--A-------------C--CA
BNL1        1d -G-NN----GT-------C--TA-G--------R-----T----------
BNL2        1d --CA---G-GT-TC----C--TA-G-------------C-----A--C-
FR2         1f -GCA---GTGT----C--A---A-G---A-T--------T--T---GGC-

HC-J6       2a -TC------G-----C--C--T--T-------------C-----TGGG-
HC-J8       2b -GCA--T--GGC---C------T-G--T-----A--TG-G------G-C-
S83         2c -TCT--T--GG-------T-----T--T---------G-G--T--CG-GC
NE92        2d ATC---T--GT-T--C--T-----G---A-A--A-----G--T--CG-G-
BNL3        2e -TC------------C--T-----G--------A--TG-G-----CG-A-
FR4         2f -TC------------C--T-----A---A-A-------------CG---
BNL4        2g -GT---T--G--------T--A------A-C------G-G--T--CG-G-
BNL5        2h -TCT--T---G----C--A--TT-G--T-----C---T-C-----CG-A-
BNL6        2i -TC------GT----C--T---T-G--T-----

NZL1        3a CGCG-----GA-G--C--T--G----------T--TA-G--T---G---
HCV-TR      3b CGCACGACAA--G--------G--G-----C------GCT-T----G---
NE48        3c T-CG--T--AT-G--------A--T-----C--T-----T------G-A-
NE274       3d AGCT--T--GT-G--C--C--G--G--T--T--C--TA-G--T--AG-C-
NE145       3e C--T-----T--G--C--C--G-----T--C--T-----T------G-C-
NE125       3f TGCA-----G--G-----A--A-----T--T--A--TT-G------G---

Z4          4a CGCG-----TT-G-----T--------T--T--------C-----AGG--
Z1          4b TGCG--T--TA-G-----C---T-----A-T--A--T--G--T--AGGC-
GB358       4c TGC---T--TGCG--C--C--T--T---A-C--A-----G-----TGGC-
DK13        4d CG-------T-----C--C---------A-C--A---G-G--T---GG--
GB809       4e TGCT------G-G--C--C----------C--C-----G-----TGGCT
BNL7        4k -GC------TG-T-----A-----T---A-C-----TT-R--T--YGGCT
BNL8        4k -GCT-----TG-T--C--A-----T---A-C-----TT-G--T--CGGCT
BNL9        4k -GCG-----TG-------A-----T---A-C-----TT-G--T--CGG--
BNL10       4k AGCT-----TG-T-----A-----T---A-C-----YT-G--T--CGGCT
BNL11       4k -GCT-----TG-T-----A-----T---A-C-----T--G-----TGGCT
BNL12       4l TGCA--T-----A-CG--T----------T--A-----C------GG--

BE95        5a AG-G--TG-------C--C--GT-A-----A--A---GCG--T---G-AC

HK2         6a CGC---AGTGG-T--C--AT----G---A-C--------G--T--C---C

FR1         7a -GCA--GG-AT-T---------A-G---A-C--A-----C--T--TAGCA

VN4         8a CGCT---G-GT-------A--TA-G--T----------G------GGCC

VN12        9a TGCT--TG-GT----C--T---A-G--------C---T-------TGGGC

NE98       10a RGCG--------A--C--A--T--------A--A-----T--T--AG-GC
```

Figure 1 - continued

```
                 851                                                    900
HCV-1       1a TCTTTCTTGTCGGCCAACTGTTCACCTTCTCTCCCAGGCGCCACTGGACG
HCV-J       1b -T-----C---TC---G-------------A--TC-C--GT-TGA----
HC-G9       1c ----C-----T-----GA-C--------------------A---T------
BNL1        1d ----C--C-CT-----G--A--------T--A---C-CATG---CAT--A
BNL2        1d ----C----------G--A--------T--A---C-CTTGT--CAT--A
FR2         1f ----C--C--T--G---T----------A-GT--C----G--T------

HC-J6       2a -GA-G----CA-C---GA------TTG----G--ACA--A-------TTT
HC-J8       2b -GA-GA--C-ATCG--GGCT----TGG-A--A--ACAA------AACTTC
S83         2c -GA-G--G-C--CT--GG-CG--GT-G-G--G--ACAA-A---TAC-TTT
NE92        2d -GA-GT-G-CTTCT---G-C----T-A----G---CA--AT--TAA-TTT
BNL3        2e -GA-GA-A-CT-CA--GGCT----T-G-GG-A--G-A------T-ACTTC
FR4         2f -GA-GA-A-CA-CG---G-TGC-GT-G----A--GCAATA---TACTTTT
BNL4        2g -GA-GA-A-CT-CT--GG-TG---TTG----G--GCAA-AT---AACTTT
BNL5        2h -GA-GT-G---TCT---T-T----TGA----C--TCA--A----ATCTTC

NZL1        3a -------C--G--A---GCC-----G---AGA--TC-A-----TCAA---
HCV-TR      3b -G--------G--A---GC----------AGA--TC-C------AC---C
NE48        3c -T--C--C--A--A---GCA-----A---AGA---C-A------CA---A
NE274       3d ----CT-G--G--A--GGCT---------AGA--TC-T-AG---AAC---
NE145       3e ----C-----G--G--GGCC--T--A---AGG--TC-T--T--TAC---T
NE125       3f -T--C-----G------GC------T---AGAG-TC---AA--T-AT--C

Z4          4a C---C--GA-G--G--GA--A----T--TCGG--GC-T----------C
Z1          4b ----C--A--G-----G------GA----CGA--GC-C--G--------C
GB358       4c -A---T-G--T--T--GA-----T-T---CAG--GC-------------T
DK13        4d -G--CT-G-----T--------------CAA--TC-C----------C
GB809       4e -A--CT-G--A------A----------CAA--GC-A----------
BNL7        4k -G--C--A-----T--GA-----T-T---CGA--A--------------T
BNL8        4k -G--CT-G--T--T--GA----TT-T---CGA--AC-A----------T
BNL9        4k CG--CT-G--T--T--GA-----T-T---CGA--AC-------------C
BNL10       4k -G--CT-G--T--T--GA-----T-T--YCAG--TC-------------T
BNL11       4k -G--C--G--T--T--GA-----T-T---CGA--AC-------------T
BNL12       4l C---C--A--G--G--GA----------CAG--GC-T----------T

BE95        5a -A--CT-G--A------A---------ATAGG--TC-C-AG---GCT---

HK2         6a -----T-G-CG--A-----A--------TCAG---C-C--T--T-----T

FR1         7a -AA-CT-G--A--G--G--T--T--T---AGG--T-A-TA---TCA-GTT

VN4         8a -T--C--C--T--A--G--C-----GC--AGG--TC--ATG--TCA-GTT

VN12        9a -------C--T--G--GT-------G---AGA------ATGT-TGA--TC

NE98       10a -A--------Y--G--GGG----T-A-GGAGA-ATC-C-AG--T-----T
```

Figure 1 - continued

```
                 901                                                  950
HCV-1       1a ACGCAAGGTTGCAATTGCTCTATCTATCCCGGCCATATAACGGGTCACCG
HCV-J       1b GTA----A----------A--------------CG--T-A--------
HC-G9       1c -------AC----------C-----C--A------G-G--A-----T--
BNL1        1d -----G-AG-----C-----A---
BNL2        1d --A--G-AG-----C-----A---
FR2         1f GT---G-AC--T-----T--C--T--CT-T-----C--------C-----

HC-J6       2a GT-----AC----------C-----C--T--TACC--C--T--A-----
HC-J8       2b --C----AG-----C--T--C-----C-AA--T--C--C--C--C--T--
S83         2c GTC--G-AA-----C--T--C--A--C--G----GC--T-----A-----
NE92        2d GTC--G-AC-----C--T--C--A--C--A-----C--C--T--A--T--
BNL3        2e GTC--G-AA----------C--A--C--A-----C--T--A-----T--
FR4         2f GTC--G-AA-----C-----C--A--C--A--------C--A--A--T--
BNL4        2g T-C--G-A---------T--C---
BNL5        2h GTC--G-A------C-----G--A

NZL1        3a GTC--GACC--T--C-----GC-G--C--A------C-TT-A--A--T--
HCV-TR      3b GT---GACG-----C-----G--A--C--A------G-TT-A--A--T--
NE48        3c GTT--GCA------C-----AC-G--C--A--T---G-TT-A-----T--
NE274       3d GT---GACC----------AC-G--C--T--T--C---T-A--A---A-
NE145       3e GTC--GACC-----C-----GT-G--C--A--------C--A--A--T--
NE125       3f GTC--GTTG----------AC-A--C--A--A--C--T--A--A--T-A

Z4          4a -----G-AG--------T--C-----CA-T--------C--C--C---A-
Z1          4b --C--G-A------C-----C--------T--T--CG-CT----C---A-
GB358       4c -----G-AC--------T--C-----CG-G--G--CG-T-----C---A-
DK13        4d --C----AC--------T--C-----CA-A--A-----C--A--A---A-
GB809       4e --C--G-AC--T-----T--C-----CG-A--G-----T-----C--T--
BNL7        4k --T----A---------T--C---
BNL8        4k G-C--G-A---------T------
BNL9        4k --C----A------C-----C---
BNL10       4k --C--G-A---------T--C---
BNL11       4k --C--G-AA--------T--C---
BNL12       4l GTC----AC-----C--T--C---

BE95        5a GT---GAAC-----C--T--C--T--CAGT------G-T--C--C-----

HK2         6a GT-----AC-----C-----C------A-A-----CG-C--C--C---A-

FR1         7a --C--G-A---T--C--------NA-CN-T-----CG-C-----A---A-

VN4         8a GTC--G-AG--T--C--T--C-----CA-A--G-----C--T--A-----

VN12        9a G-C--G-AC-----C--T--C------G-A-----C--C--T--G-----

NE98       10a GTC--G-AC-----C--T--C---
```

Figure 1 -continued

		951 957
HCV-1	1a	CATGGCA
HCV-J	1b	------T
HC-G9	1c	A-----T
FR2	1f	NNNNNNN
HC-J6	2a	------G
HC-J8	2b	-------
S83	2c	------T
NE92	2d	G-----G
BNL3	2e	------G
FR4	2f	A----NN
NZL1	3a	A-----T
HCV-TR	3b	T-----G
NE48	3c	G-----T
NE274	3d	G-----T
NE145	3e	-------
NE125	3f	T-----T
Z4	4a	G-----G
Z1	4b	G-----C
GB358	4c	G------
DK13	4d	A-----T
GB809	4e	G-----T
BE95	5a	G------
HK2	6a	G-----T
FR1	7a	G------
VN4	8a	A------
VN12	9a	G-----G

Figure 2

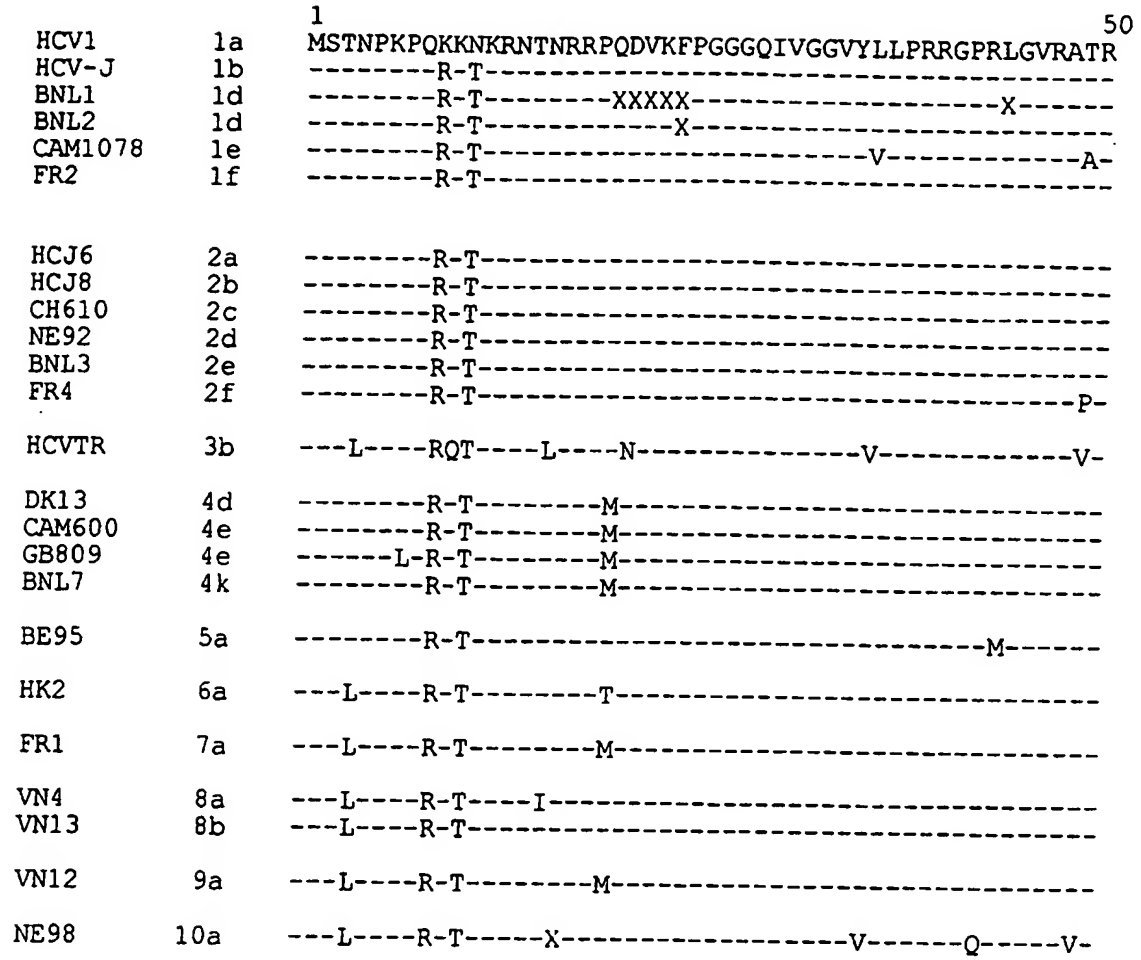

1
50
HCV1 1a MSTNPKPQKKNKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATR
HCV-J 1b --------R-T----------
BNL1 1d --------R-T--------XXXXX----------X------
BNL2 1d --------R-T----------X----------
CAM1078 1e --------R-T----------V----------A-
FR2 1f --------R-T----------
HCJ6 2a --------R-T----------
HCJ8 2b --------R-T----------
CH610 2c --------R-T----------
NE92 2d --------R-T----------
BNL3 2e --------R-T----------
FR4 2f --------R-T----------P-
HCVTR 3b ---L----RQT----L----N----------V----------V-
DK13 4d --------R-T--------M----------
CAM600 4e --------R-T--------M----------
GB809 4e ------L-R-T--------M----------
BNL7 4k --------R-T--------M----------
BE95 5a --------R-T----------M------
HK2 6a ---L----R-T--------T----------
FR1 7a ---L----R-T--------M----------
VN4 8a ---L----R-T----I----------
VN13 8b ---L----R-T----------
VN12 9a ---L----R-T--------M----------
NE98 10a ---L----R-T-----X----------V------Q-----V-

Figure 2 - continued

```
                    51                                                  100
HCV1        1a      KTSERSQPRGRRQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSP
HCV-J       1b      ----------------------------------------M---------
BNL1        1d      ------------------X-X---S-------------------X-----
BNL2        1d      ---------D---------QSD-XX-----H-------------------
CAM1078     1e      ----------------E------
FR2         1f      ------------------------S-----------A-------------

HCJ6        2a      ----------------D--ST-KS-GK------------L---------
HCJ8        2b      ----------------D--ST-KS-GK----------------------
CH610       2c      ----------------D--TT-KS-GR------------L---------
NE92        2d      ----------------D---T-KS-GK------------L---------
BNL3        2e      ----------------D-XAT--S-GR------------L---------
FR4         2f      ----------------D--AT-KS-GR------------L---------

HCVTR       3b      ---------KQ-HL-----SR---S-------------K---L-------

DK13        4d      ------------------QL---S-------------------------
CAM600      4e      ------------------T---S-------------------------
GB809       4e      ------------------S---S-------------------------
BNL7        4k      ------------------S---S-------------------X-----

BE95        5a      ------------------Q-T--S-G---------A---L---------

HK2         6a      ------------------Q-Q--H-------------------------

FR1         7a      ----------------V-Q-T--S-G-----------------------

VN4         8a      ----------------V-HQT---------------------------
VN13        8b      ----------------V-HQT---------------------------

VN12        9a      -------A---------V-QNQ---------------------------

NE98       10a      ---------S------R---T---S-------------------------
```

Figure 2 - continued

```
                  101                                                                 150
HCV1        1a    RGSRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARA
HCV-J       1b    --------------------------------------------------
BNL1        1d    -----N---
BNL2        1d    ---------
FR2         1f    ---------N-----------------------------------------S-T

HC-J6       2a    ---------N---H----V-------------------V-------V---
HC-J8       2b    -----T-------H------R----I------------V----V--V---
CH610       2c    -------------H-------------------------V----V--V---
NE92        2d    -------------H-------------------------V----V--V---
BNL3        2e    --------                        --XX------X-V----V--X---
FR4         2f    ---------N---H----------------X-------V----V--V---

HCV-TR      3b    ---------N--------F-----------------------V--V---

GB116       4c                                  ----------------V--V---
DK13        4d    ---------N-----------------------------V----V--V---
CAM600      4e    -X--X----N---X-------------------------------V--V---
GB809       4e    ---------N-----------------------------------V--V---
G22         4f                                  ----------------V--V---
GB549       4g                                  ----------------V--V---
GB438       4h                                  ----------------V--V---
BNL7        4k    ---------N-------

BE95        5a    -----N---N----K--------------------------G-I--V---

HK2         6a    -----H---N-------------------------------V-------V-A-

FR1         7a    -----N---N----------------XXL--------VL-G----V-A-

VN4         8a    -----N---N-------------------------------V----X--V-X-
VN13        8b    X----N---N---X-------------XX----IE--

VN12        9a    -----D-X-N---X--------------------E---V-------V-AE

NE98       10a    ---------N-----
```

Figure 2 - continued

```
                    151                                                                 200
HCV1        1a      LAHGVRVLEDGVNYATGNLPGCSFSIFLLALLSCLTVPASAYQVRNSTGL
HCV-J       1b      ------------------------------------I-----E---VS-I
BNL1        1d      ----------------------------------XT-HE---AS-V
BNL2        1d      ---------------------F-------TT-HE---AS-V
FR2         1f      -X------XG--XXXXX--X---XX----X--------T---E-HST-DG

HC-J6       2a      ------------F--------------------I-T-V--AE-K-ISTG
HC-J8       2b      -----------I----------------V---V--VE---ISSS
CH610       2c      -----------I---------------S-----IS--V--VE-K-TSTS
NE92        2d      -----------I----------------I---V-GL--K-TSSS
BNL3        2e      --X--------I--X-----------X------V---V-XVE-K-TSQA
FR4         2f      -----------I----------------I---V--I--K-NSHF
BNL4        2g      ----------------------V---V--V--K-TSTM
BNL5        2h      --I----------------------V--K-TSHS
BNL6        2i      --I----------------I---V--V--A-RS-S

HCV-TR      3b      ------A-G----------------F----C---GLEYT-TS--

GB116       4c      -E----AV---I-------------S----------T--VNY--AS-V
DK13        4d      ------L-----------------------------NY---S-V
CAM600      4e      ------AV---I----------------------T--VNY--AS-I
GB809       4e      ------AV---I--------------------GVNY--AS-V
G22         4f      ------AV---I-----------------VHYH-TS-I
GB549       4g      ------AV---I-----------------QHY--IS-I
GB438       4h      ------AV---I----------V---R-------QHY--AS-I
BNL7        4k      --I-F----------------INY--VS-I
BNL8        4k      --I----------------INY--TS-I
BNL9        4k      --I----------------INYH-TS-I
BNL9        4k      --I------I---X-----X----------TNY--VS-I
BNL10       4k      --I-----X----------TNY--VS-I
BNL11       4l      --I-------------I----------QHY--VS-I

BE95        5a      ------------------------I----------VPY--AS-I

HK2         6a      ------AI---I----------T----LTYG--S--

FR1         7a      ------AI----------T----I--K-AS-I

VN4         8a      -----XXI--X-----X---XX-X--X---------T----AHYT-KS--

VN12        9a      -X----AI---I-------------X--------T----LNYA-KS--

NE98        10a             --I-F----------F---LT-TAGLEY--AS--
```

Figure 2 - continued

```
                    201                                                                250
HCV-1       1a      YHVTNDCPNSSIVYEAADAILHTPGCVPCVREGNASRCWVAMTPTVATRD
HCV-J       1b      -------S----------M-M----------S-F------L---L-A-N
BNL1        1d      -------S----I--MDGM-M-Y---------D-HL---M-L---L-VKX
BNL2        1d      --L----S----I--MSGM---A---------N-S----MXL---L-VK-
FR2         1f      -------S-G------K-I------X---I----I-----PL---L-A-I

HC-J6       2a      -M-----T-D--TWQLQA-V--V------EKV--T----IPVS-N--VQQ
HC-J8       2b      -YA----S-N--TWQLT--V--L------ENDNGTLH--IQV--N--VKH
CH610       2c      -M-----S-----WQLEG-V---------EQI--------PVS-N--I-Q
NE92        2d      -M-----Q-----WQLR--V--V------EEK--I----IPVS-NI-VSQ
BNL3        2e      -MA----S-N---WQLX--V--V------ENSSGRFH--IPIS-NI-VSK
FR4         2f      -MA----A-D---WQLR--V--V------E-S--RTF--T-VS-N--VSR
BNL4        2g      -MA----S-N--IWQMQG-V--V------ELQ--K----IPV--N--VNQ
BNL5        2h      -M-----S-----WQLK--V--V------E-HQ-Q----IPV--N--VSQ
BNL6        2i      -M-----S-----WQLEE-V--V------EWKD-T----IPV--NI-VSQ

HCVTR       3b      -VL----S-G------E-V---L-------TT--Q-S--TTVST---V-T

GB116       4c      --I-------------DYH---L---L----V--Q------L-----APY
DK13        4d      ---------------TDYH---L-----------K-T---SL-----AQH
CAM600      4e      --I------A-----TENH---L--------T--Q------L-----SPY
GB809       4e      --I------A-----TDNH---L-------KT--Q------L-----SPY
G22         4f      --L----------F--VHH---L--------T--Q------L---L-APY
GB549       4g      ----------------DHH-M-L--------T--T-----PL-----APY
GB438       4h      ----------------DHH-M-L--------T--V----IPL-----VPY
BNL7        4k      -Y--------------DHH---L----------Q------L-----APY
BNL8        4k      ----------------DHH---L--------T--Q------L-----APY
BNL9        4k      --I-------------DHH---L--------V--Q-S----L---I-APY
BNL9        4k      ----------------DHH--AL--------V--Q------L-----APY
BNL10       4k      -------------F--DHH---L-------K---H------L-----APY
BNL11       4l      ---------------SDHH---L-------KT--T------L-----API
GB724       4x      --I--------V---TDHH---L--------T--V----TPV-----AVS

BE95        5a      ----------------DNL---A-------MT--V-----QI---LSAPS

HK2         6a      --L----------L--DAM---L---L----VDDR-T--H-V---L-IPN

FR1         7a      --L----S-N---F--ETM---L------IKA--E----LPVS--L-VPN

VN4         8a      --L-------------ETL---L-------KXX-Q-----QAS--L-VPN

VN12        9a      --L-------------NGM---L-------KT--LTK--LSAS--L-VQN

NE98       10a      -M-----S-G------G-I---L--------S--T----IPVSX---VKS
```

Figure 2 - continued

		251 300
HCV-1	1a	GKLPATQLRRHIDLLVGSATLCSALYVGDLCGSVFLVGQLFTFSPRRHWT
HCV-J	1b	SSI-T-TI---V-----A-A----M----------S---------YE-
BNL1	1d	ASV-TXAI---V-----XX-F---M--X--------A---------M-H-
BNL2	1d	ANV-TAAI---V-----T-AFR--M----------LYH-
FR2	1f	ANA-IDEV---V-----A-VF---M-I-----G----------TS-----
HC-J6	2a	PGALTQG--T---MV-M----------G-M-AA-M-IV--QH--F
HC-J8	2b	RGALTRS--T-V-MI-MA--A--------V--A-MILS-A-MV--Q--NF
CH610	2c	PGTLTKG--A-V-VI-M----------V--ALMIAA-AVIA--Q--TF
NE92	2d	PGALTKG--T---TIIA---F-----I-----A-M-AS-V-II--QH-KF
BNL3	2e	PGALTKG--AR--AV-M----------V--A-MIAA-A-IVA-K--YF
FR4	2f	PGALTRG--A---TI-M---------I-----A-MIAA-VAVV--QY-TF
BNL4	2g	PGALTRG--T---TI-MV--------I--V--A-MIAA-VVIV--QH-NF
BNL5	2h	PGALTRG--T---TI-A---V--------F--A-M--S-F-MI--QH-IF
BNL6	2i	PGAXTKG--T---II-A---F------
HCVTR	3b	LGVTTASI-T-V-M---ARQ---------AF-A------A---R----T-
GB116	4c	VGA-LES--S-V--M--A--V-----I-----G------M-S-Q------
DK13	4d	LNA-LES----V--M--G--------I--V--G----------Q------
CAM600	4e	AGA-LEP----V--M--A--M-----I-----GL-----M---Q------
GB809	4e	VGA-LEP----V--M--A--V----------GL-----M---Q------
G22	4f	LGA-LESM---V--M--T----------GI--A--M---R--L---
GB549	4g	VGA-LESM---V--M--A--V-----I-----G------M---R------
GB438	4h	LGA-L-SV-Q-V--M--A--------I--H--G---A--MVS-Q------
BNL7	4k	IGA-LES--S-V--M--A--V-----I--X-XGL-----M-S-R------
BNL8	4k	IGA-LES--S-V--M--A--V-----I-----GL-----M-S-R------
BNL9	4k	IGA-LES--S-V--M--A--V-----I-----GA-----M-S-R------
BNL9	4k	TAA-LES--S-V--M--A--V-----I-X---GL-----M-SXQ------
BNL10	4k	IGA-LES--S-V-VM--A--V-----I-----GL-----M-S-R------
BNL11	4l	LSA-LMSV---V--M--A---S----------GA-----M---Q------
GB724	4x	VDA-LESF---V--M--A-----V--------GA-----M---Q------
BE95	5a	LGAVTAP---AV-Y-A-G-A---------A--AL-----M--YR--Q-A-
HK2	6a	AST---GF---V---A-A-VV--S--I------L--A------Q------
FR1	7a	SSV-IHGF---V-----A-AF---M-I------II--------R-KY-QV
VN4	8a	AST-V-GF-K-V-IM--A-AF---M-------GL--------LR--M-QV
VN12	9a	ASVSIRGV-E-V-----A-AF---M-------GL---------R--MYEI
NE98	10a	PCAATAS--T-V-MM-XA----------AL--X--G-SWRH-Q---

Figure 2 - continued

		301 319
		TQGCNCSIYPGHITGHRMA
HCV-1	1a	
HCV-J	1b	V-D---------VS-----
BNL1	1d	--E-----
BNL2	1d	--E-----
FR2	1f	V-D------S------XXX
HC-J6	2a	V-D--------T-------
HC-J8	2b	--E------Q---------
CH610	2c	V-E---------------X
NE92	2d	V-D----------------
BNL3	2e	V-E----------------
FR4	2f	V-E---------------X
BNL4	2g	S-D-----
BNL5	2h	V-D-----
HCVTR	3b	V-T---------VS-----
GB116	4c	--D------A--V------
DK13	4d	--D------T---------
CAM600	4e	--D------T---------
GB809	4e	--D------A---------
G22	4f	--E----T----------
GB549	4g	--D------D---------
GB438	4h	--D------V---------
BNL7	4k	--D-----
BNL8	4k	A-D-----
BNL9	4k	--D-----
BNL9	4k	--D-----
BNL10	4k	--E-----
BNL11	4l	V-D-----
GB724	4x	--D------T---------
BE95	5a	V-N------S--V------
HK2	6a	V-D------T--V------
FR1	7a	--D----XNX--V------
VN4	8a	V-E------T---------
VN12	9a	A-D------A---------
NE98	10a	V-D-----

Figure 3

SEQ ID NO. 1 (BNL1, 1d)
ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCTCAKGGSGTN
NNNNNNCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGNNG
GGTGTGCGCGCGACTAGGAAGACTTCCGAGCGGTCACAACCTCGTGGCAGGCGACAGCCTATCCCC
AAGGCTCGYCGGYCCGAGGGCAGGTCCTGGGCTCAGCCCGGGTATCCTTGGCCCCTCTATGGCAAT
GAGGGCTGCGGGTGGGCGGGNTGGCTCCTGTCCCCCCGCGGCTCTCGGCCCAATTGGGGCCCC

SEQ ID NO. 3 (BNL1, 1d)
GACGGCGTGAACTATGCAACAGGGAACTTGCCCGGTTGCTCTTTCTCTATCTTCCTCTTGGCTTTG
CTGTCCTGCTTGACGGTTCCAACKACCGCTCACGAGGTGCGCAACGCATCCGGGGGTGTATCATGTC
ACCAACGACTGTTCCAACTCGAGCATCATCTATGAGATGGACGGTATGATCATGCACTACCCAGGG
TGCGTGCCCTGCGTTCGGGAGGATAACCATCTCCGCTGCTGGATGGCGCTCACCCCCACGCTTGCG
GTCAAAAAYGCTAGTGTCCCCACTRCGGCAATCCGACGTCACGTCGACTTGCTTGTTGGGGGNNCC
ACGTTCTGTTCCGCTATGTACGTGGGRGACCTTTGCGGGTCTGTCTTCCTCGCTGGCCAGCTATTC
ACCTTTTCACCCCGCATGCACCATACAACGCAGGAGTGCAACTGCTCAATC

SEQ ID NO. 5 (BNL2, 1d)
ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCACAGGACGTC
AAGNTCCCGGGTGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCCGCGCAGGGGCCCCAGGTTG
GGTGTGCGCGCGACCAGGAAGACTTCCGAGCGGTCGCAGCCTCGTGACAGGCGACAGCCTATTCCT
AAGGCTCGCCAGTCCGATGGCAGNNCCTGGGCTCAGCCAGGGCATCCCTGGCCCCTCTATGGCAAT
GAGGGCTGCGGATGGGCGGGATGGCTCCTGTCCCCCCGCGGCTCTCGGCCCAGTTGGGGCCCC

SEQ ID NO. 7 (BNL2, 1d)
GACGGCGTGAACTATGCAACAGGGAATTTGCCTGGTTGCTCTTTCTCTATCTTCCTCTTAGCTTTT
CTGTCCTGCTTGACGGTTCCAACTACCGCTCATGAGGTGCGCAACGCATCCGGGGTATATCATCTC
ACCAATGACTGTTCCAACTCGAGCATCATCTATGAGATGAGTGGTATGATCTTGCACGCCCCAGGG
TGTGTGCCCTGCGTTCGGGAGAACAACTCTTCTCGTTGCTGGATGCCRCTCACCCCCACGCTTGCG
GTCAAAGACGCTAATGTCCCTACTGCGGCAATCCGACGCCATGTCGACTTGCTGGTTGGGACAGCC
GCGTTTCGTTCCGCTATGTACGTGGGGGACCTCTGCGGATCCGTCTTCCTTGTCGGCCAGCTATTC
ACCTTTTCACCCCGCTTGTACCATACAACACAGGAGTGCAACTGCTCAATC

SEQ ID NO. 9 (CAM1078, 1e)
ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGCCGCCCACAGGACGTC
AAGTTCCCGGGCGGTGGCCAGATCGTTGGTGGAGTCTACGTGCTACCGCGCAGGGGCCCTAGATTG
GGTGTGCGCGCAGCGCGGAAGACTTCGGAGCGGTCGCAACCTCGTGGGAGGCGCCAACCTATTCCC
AAGGAGCGCCGACCCGAGGGCAGGT

Figure 3 - continued

SEQ ID NO. 11 (FR2, 1f)
ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGCAACACCAACCGCCGCCCACAGGACGTT
AAATTCCCGGGTGGGGGGCAGATCGTGGGTGGAGTTTACTTGTTGCCGCGCAGGGGCCCCAGGTTG
GGTGTGCGCGCGACGAGGAAGACTTCCGAGCGGTCGCAACCTCGCGGAAGGC
GACAGCCTATCCCCAAGGCTCGCCGACCCGAGGGCAGGTCCTGGGCTCAGCCTGGGTACC
CATGGCCCCTCTATGCTAACGAGGGCTGCGGATGGGCGGGATGGCTCCTGTCCCCTCGCG
GCTCCCGTCCTAGCTGGGGCCCCAATGACCCCCGACGTAGATCACGCAATTTGGGTAAGG
TCATCGATACCCTAACGTGTGGCTTCGCCGATCTCATGGGGTACATTCCGCTCGTCGGCGC
CCCCCTAGGGGGCGCTTCCAGAACCCTGNCACATGGTGTCCGGGTCCTGGNAGGCGGCGTGATNNN
NNNNNNNNNNAACCTTCCNGGTTGCTCTTTNNCTATCTTCCTCTTGGCNTTACTCTCTTGCCTCAC
AGTCCCCACCTCTGCCTATGAGGTGCACAGCACAACCGATGGCTACCATGTCACTAATGACTGTTC
CAACGGCAGCATCGTATATGAGGCAAAGGACATCATCCTTCACACGCCTGGGTGNGTGCCCTGCAT
ACGGGAAGGCAATATCTCCCGTTGCTGGGTACCGCTCACCCCCACGCTCGCAGCGCGGATCGCGAA
CGCTCCCATCGATGAGGTGCGGCGTCACGTCGACCTCCTCGTGGGGGCAGCCGTGTTCTGCTCAGC
CATGTACATTGGGGACCTTTGTGGGGGCGTCTTCCTCGTTGGGCAATTGTTCACCTTCACGTCCCG
GCGGCATTGGACGGTGCAGGACTGTAATTGTTCCATTTACTCTGGCCACATAACGGGCCACCGNNN
NNNN

SEQ ID NO. 13 (BNL3, 2e)
ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAATACCAACCGCCGCCCACAGGACGTC
AAGTTCCCGGGCGGCGGCCAGATCGTTGGCGGAGTTTACTTGTTGCCGCGCAGGGGCCCCAGATTG
GGTGTGCGCGCGACGAGAAAGACTTCTGAACGGTCCCAGCCACGTGGAAGGCGCCAGCCCATCCCT
AAAGATCGGNGNGCCACTGGCAGGTCCTGGGGACGTCCAGGATATCCCTGGCCCCTGTATGGGAAC
GAGGGGCTCGGCTGGGCAGGATGGCTCCTGTCCCCCCGAGGCTCTC

SEQ ID NO. 15 (BNL3, 2e)
ACGTGCGGNTNTGCCGACCTCATGGGGTACATNCCCGTTGTCGGCGCCCCGGTGGGCGGGGTNGC
CAGGGCCCTCGCGNATGGCGTGCGGGTCCTGGAGGACGGGATAAATTATGNAACAGGGAACCTCCC
TGGTTGCTCCTTTTCTATCTTCTNGTTGGCTCTTCTGTCTTGTGTCACCGTGCCTGTCTCTGNCGT
TGAGGTCAAAAATACCAGTCAGGCCTATATGGCAACCAACGACTGCTCCAACAACAGCATCGTATG
GCAATTGGNGGACGCGGTGCTTCATGTTCCTGGATGTGTCCCCTGCGAGAATAGCTCCGGTCGGTT
CCACTGTTGGATCCCGATCTCGCCCAACATAGCCGTGAGCAAACCTGGTGCTCTCACCAAGGGACT
GCGGGCACGCATTGATGCCGTCGTGATGTCCGCCACCCTCTGCTCTGCCCTGTACGTGGGAGATGT
GTGCGGCGCAGTGATGATAGCTGCACAGGCTTTCATCGTGGCACCGAAGCGCCATTACTTCGTCCA
GGAATGCAATTGCTCCATATACCCAGGCCACATTACAGGTCATCGCATGGCG

SEQ ID NO. 17 (FR4, 2f)
ATGAGCACAAATCCTAAACCTCAAAGAAAAACTAAAAGAAACACTAACCGTCGCCCACAGGAC
GTTAAGTTCCCGGGCGGCGGCCAGATCGTTGGCGGAGTTTACTTGTTGCCGCGCAGGGGCCCCAG
GTTGGGTGTGCGCGCGCCAAGGAAGACTTCTGAACGGTCCCAGCCACGTGGAAGGCGCCAGCCC
ATCCCAAAAGATCGGCGCGCCACTGGCAAGTCCTGGGGACGTCCAGGATACCCTTGGCCCCTGT
ACGGGAACGAGGGCCTCGGCTGGGCAGGGTGGCTCCTGTCCCCCCGGGGCTCTCGCCCCTCGTG
GGGCCCAAACGACCCCCGGCACAGGTCACGCAACTTGGGTAAGGTCATCGATACCCTCACGTG
TGGCTTTGSCGACCTCATGGGGTACATACCTGTCGTCGGCGCCCCTGTGGGCGGCGTTGCCAGA
GCCCTCGCGCATGGCGTGCGGGTCCTGGAGGACGGGATAAATTATGCAACAGGGAACTTGCCCGGT
TGCTCCTTTTCTATCTTCTTGCTGGCTCTCTTGTCTTGTATCACCGTGCCCGTGTCTGCCATACAG
GTTAAGAACAACAGCCACTTCTACATGGCGACTAATGACTGTGCCAATGACAGCATCGTCTGGCAG
CTCAGGGACGCGGTGCTCCATGTTCCTGGATGTGTCCCCTGTGAGAGGTCAGGTAATAGGACCTTC
TGTTGGACAGCGGTCTCGCCCAACGTGGCTGTGAGCCGACCTGGTGCTCTCACTAGAGGTCTGCGG
GCTCACATTGATACCATCGTGATGTCCGCCACCCTCTGCTCTGCCCTATACATAGGGGACCTATGC
GGCGCTGTGATGATAGCAGCGCAAGTTGCCGTCGTCTCACCGCAATACCATACTTTTGTCCAGGAA
TGCAACTGCTCCATATACCCAGGCCATATCACAGGACATCGAATGGNN

Figure 3 - continued

SEQ ID NO. 19 (BNL4, 2g)
GACGGGGTAAATTATGCAACAGGGAATCTGCCTGGTTGCTCTTTCTCTATCTTCTTGTTGGCTCTT
CTGTCTTGTGTCACCGTGCCTGTCTCTGCCGTGCAGGTTAAGAACACCAGTACCATGTACATGGCA
ACCAATGACTGTTCCAACAACAGCATCATCTGGCAAATGCAGGGCGCGGTGCTTCATGTTCCTGGA
TGTGTCCCGTGTGAGTTGCAGGGCAATAAGTCCCGGTGCTGGATACCGGTCACTCCCAACGTGGCT
GTGAACCAGCCCGGCGCCCTCACTAGGGGCTTGCGGACGCACATTGACACCATCGTGATGGTCGCT
ACGCTCTGTTCTGCACTCTACATCGGGGACGTGTGTGGCGCGGTGATGATAGCTGCTCAGGTTGTC
ATTGTCTCGCCGCAACATCACAACTTTTCCCAGGATTGCAATTGTTCCATC

SEQ ID NO. 21 (BNL5, 2h)
ATGAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACTAACCGCCGCCCACAGGACGTT
AAGTTCCCGGGCGGTGGCCAGATCGTTGGCGGAGTATACTTGTTGCCGCGCAGGGGCCCCCGGTTG
GGTGTGCGCGCGACGAGGAAAACTTCCGAACGGTCCCAGCCACGTGGGAGGCGCCAGCCCATCCCT
AAAGATCGGCGCTCCACTGGCAAATCCTGGGGACGTCCAGGATACCCTTGGCCCCTGTATGGGAAC
GAGGGCCTTGGTTGGGCAGGATGGCTCTTGTCCCCTCGAGGCTCTC

SEQ ID NO. 23 (BNL5, 2h)
GACGGGATAAACTACGCAACAGGGAATCTGCCCGGTTGCTCCTTTTCTATCTTCTTGCTGGCCTTG
CTATCCTGTCTCACTGTGCCGGCGTCCGCTGTGCAGGTCAAGAACACCAGCCACTCTTATATGGTG
ACCAATGATTGCTCAAACAGCAGCATTGTCTGGCAGCTTAAGGATGCTGTGCTTCACGTCCCTGGA
TGTGTTCCATGTGAGAGGCACCAAAATCAGTCTCGCTGCTGGATACCTGTGACACCCAATGTGGCC
GTGAGCCAACCTGGCGCGCTCACCAGGGGTTTGCGGACGCACATTGACACCATCGTTGCGTCTGCT
ACCGTCTGCTCAGCTTTGTATGTGGGCGACTTCTGCGGCGCAGTGATGTTGGTCTCTCAATTTTTC
ATGATCTCCCCTCAGCACCACATCTTCGTCCAGGATTGCAACTGCTCGATA

SEQ ID NO. 25 (BNL6, 2i)
GACGGGATAAACTATGCAACAGGGAACCTGCCTGGTTGCTCCTTTTCTATCTTCTTACTGGCCCTG
CTTTCTTGCATCACCGTGCCGGTCTCTGCCGTGCAAGTTGCGAACCGCAGTGGTTCTTACATGGTG
ACCAATGATTGCTCGAACAGCAGCATCGTTTGGCAGCTCGAGGAGGCCGTCCTTCACGTCCCTGGA
TGTGTTCCCTGTGAGTGGAAGGACAACACCTCCCGCTGCTGGATACCGGTCACCCCTAACATCGCT
GTGAGCCAACCTGGCGCGCTTACCAAGGGCCTGCGGACACATATTGACATCATTGTCGCGTCCGCC
ACGTTCTGCTCTGCCTTGTATGTGGG

SEQ ID NO. 27 (BNL7, 4k)
ATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACACCAACCGCCGCCCCATGGACGTT
AAGTTCCCGGGTGGTGGCCAGATCGTTGGCGGAGTTTACTTGTTGCCGCGCAGGGGCCCCAGGTTG
GGTGTGCGCGCGACTCGGAAGACTTCGGAGCGGTCGCAACCTCGTGGGAGACGCCAACCTATCCCC
AAGGCGCGTCGATCCGAGGGAAGGTCCTGGGCACAGCCAGGATATCCATGGCCTCTTTACGGTAAT
GAGGGTTGCGGGTGGGCANNATGGCTCTTGTCCCCCCGCGGTTCTC

SEQ ID NO. 29 (BNL7, 4k)
GACGGGATCAATTTTGCAACAGGGAACCTCCCCGGTTGCTCCTTTTCTATCTTCCTCTTGGCACTC
CTCTCGTGCCTGACTGTCCCCGCTTCGGCCATCAACTATCGCAATGTCTCGGGCATTTACTATGTC
ACCAATGATTGCCCGAATTCAAGCATAGTGTATGAGGCCGACCATCACATCTTGCACCTCCCAGGT
TGCGTGCCCTGCGTGAGAGAGGGGAATCAGTCACGTTGCTGGGTAGCCCTTACCCCTACCGTCGCA
GCGCCATACATCGGCGCGCCACTTGAGTCTCTACGGAGTCATGTGGACTTGATGGTGGGGGCCGCC
ACTGTTTGTTCAGCCCTTTACATCGGGGATTTRTGTGGYGGCTTGTTCCTAGTCGGTCAGATGTTC
TCTTTCCGACCAAGGCGCCACTGGACTACTCAAGATTGCAATTGTTCCATC

Figure 3 - continued

SEQ ID NO 31 (BNL8, 4k)
GACGGGATCAATTATGCAACAGGGAACCTTCCCGGTTGCTCTTTTTCTATCTTCCTCTTGGCACTC
CTCTCGTGCCTGACTGTTCCCGCTTCGGCCATTAACTACCGCAACACCTCGGGCATCTACCACGTC
ACCAATGACTGCCCGAACTCGAGCATAGTTTATGAGGCCGACCACCACATCTTGCACCTTCCAGGT
TGCGTGCCCTGCGTGAGAACTGGGAATCAGTCACGTTGCTGGGTGGCCCTTACTCCTACCGTCGCA
GCGCCATACATCGGCGCACCGCTTGAGTCTCTGCGGAGTCATGTGGATCTGATGGTGGGGGCTGCC
ACTGTTTGCTCAGCCCTTTACATCGGGGATTTGTGTGGCGGCTTGTTCTTGGTTGGTCAGATGTTT
TCTTTCCGACCACGACGCCACTGGACTGCCCAGGATTGCAATTGTTCTATC

SEQ ID NO. 33 (BNL9, 4k)
GACGGGATTAATTATGCAACAGGGAATCTTCCCGGTTGCTCCTTTTCTATCTTCCTCTTGGCACTT
CTCTCGTGCCTGACTGTCCCCGCTTCGGCCATTAACTACCACAACACCTCGGGCATCTATCATATC
ACCAACGACTGCCCGAATTCAAGCATAGTGTATGAGGCCGACCATCACATCTTGCATCTCCCAGGT
TGCGTGCCCTGCGTGAGAGTGGGGAATCAGTCGAGTTGCTGGGTGGCCCTTACCCCTACCATCGCA
GCGCCATACATCGGCGCACCGCTTGAGTCCTTGCGGAGTCATGTGGATCTGATGGTGGGGGCGGCC
ACTGTCTGTTCAGCCCTTTACATCGGGGATTTGTGTGGCGGTGCGTTCTTGGTTGGTCAGATGTTC
TCTTTCCGACCACGGCGCCACTGGACCACCCAAGATTGCAACTGCTCCATC

SEQ ID NO. 35 (BNL10, 4k)
GACGGGATCAATTATGCAACAGGGAATATTCCCGGTTGCTCYTTTTCTATCTTCCTTYTGGCACTT
CTCTCGTGTCTGACTGTCCCCGCTTCGGCCACTAACTATCGCAACGTCTCGGGCATCTACCATGTC
ACCAATGACTGCCCGAATTCAAGCATAGTGTATGAGGCCGACCATCACATCTTAGCACTTCCAGGT
TGCGTGCCCTGCGTGAGAGTGGGGAACCAGTCACGCTGCTGGGTGGCCCTTACCCCTACCGTCGCA
GCGCCATACACCGCGGCGCCGCTTGAGTCCCTGCGGAGTCATGTGGATCTGATGGTGGGAGCTGCC
ACTGTTTGTTCAGCCCTTTACATCGGGGAYTTGTGTGGCGGCTTGTTCTTGGTTGGTCAGATGTTC
TCTTTYCAGCCTCGGCGCCACTGGACTACCCAGGATTGCAATTGTTCCATC

SEQ ID NO. 37 (BNL11, 4k)
GACGGGATTAATTATGCAACAGGGAAYCTCCCCGGTTGCTCTTTTTCTATCTTCCTCTTGGCACTT
CTCTCGTGCCTGACTGTCCCCGCTTCGGCCACCAACTACCGCAATGTCTCGGGCATTTACCATGTC
ACCAATGACTGCCCGAATTCAAGCATAGTGTTTGAGGCCGACCATCACATCTTGCACCTTCCAGGA
TGCGTGCCCTGCGTGAAAGAGGGAAATCATTCACGCTGCTGGGTGGCCCTTACCCCTACCGTCGCA
GCGCCATACATCGGCGCGCCACTTGAGTCTCTACGGAGTCATGTGGATGTGATGGTGGGGGCTGCC
ACTGTTTGTTCAGCCCTTTACATCGGGGATCTGTGCGGTGGCTTGTTCCTGGTTGGTCAGATGTTC
TCTTTCCGACCACGGCGCCACTGGACTACCCAGGAATGCAATTGTTCCATC

SEQ ID NO. 39 (BNL12, 41)
GACGGGATCAATTATGCAACAGGGAACCTCCCCGGTTGCTCTTTCTCTATCTTCATCCTGGCACTT
CTCTCGTGCCTGACTGTCCCGGCCTCGGCTCAGCATTATCGGAATGTCTCGGGCATTTACCACGTC
ACCAACGACTGCCCGAACTCCAGCATAGTGTATGAGTCCGACCATCACATCTTACACCTACCAGGG
TGTGTACCCTGTGTGAAGACTGGGAACACTTCGCGCTGCTGGGTGGCCTTAACACCTACCGTGGCC
GCGCCCATACTTTCGGCTCCACTTATGTCCGTACGGCGGCATGTGGATCTGATGGTGGGTGCAGCT
ACCCTATCGTCTGCCCTCTACGTTGGAGACCTCTGCGGGGGTGCCTTCCTAGTGGGGCAGATGTTC
ACCTTCCAGCCGCGTCGCCACTGGACTGTCCAAGACTGCAACTGTTCCATC

SEQ ID NO. 45 (VN13, 7a)
ATGAGCACACTTCCTAAACCTCAAAGAAAAACCAAACGAAACACCAACCGTCGCCCACAGGACGTC
AAGTTCCCGGGTGGCGGTCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGGCCCTCGTTTG
GGTGTGCGCGCGACGAGGAAAACTTCTGAACGGTCCCAGCCCAGGGGTAGACGCCAACCTATACCG
AAGGTGCGTCACCAAACGGGCCGTACCTGGGCTCAACCCGGGTACCCCTGGCCTCTTTATGGGAAT
GAGGGTTGTGGCTGGGCAGGGTGGCTCCTGTCCCCCCNCGGCTCTCGCCCTAATTGGGGCCCTAAT
GACCCCCGGNGGAGGTCCCGCAACCTGGGTAAGGTCATCGATACCCTTACTTGNGGSTTCGCCGAC
CTCATAGAGTACATTCC

Figure 3 - continued

SEQ ID NO. 43 (VN4, 7c)
ATGAGCACACTTCCAAAACCCCAAAGAAAAACCAAAAGAAACACCATCCGCCGCCCACA
GGACGTCAAGTTCCCGGGTGGCGGCCAGATCGTTGGTGGAGTCTACTTGCTGCCGCGCAG
GGGCCCGCGCTTGGGTGTGCGCGCGACGAGAAAGACTTCTGAACGGTCCCAGCCCAGAGG
TAGGCGCCAACCAATACCCAAAGTGCGCCACCAAACGGGCCGTACCTGGGCCCAGCCCGG
GTACCCCTGGCCTCTTTATGGAAATGAGGGCTGTGGTTGGGCAGGCTGGCTCCTGTCCCC
CCGCGGCTCTCGCCCAAATTGGGGCCCAAACGACCCCCGGCGGAGGTCCCGCAACTTGGG
TAAAGTCATCGACACCCTTACTTGCGGCTTCGCCGACCTCATGGGGTATATCCCTGTCGTAG
GCGCTCCGWTGGGAGGCGTCGCGGNGGCCTTGGCGCATGGGGTCANGGNCATCGAGGACGGNGTAA
ATTACGCAACAGNGAATCTTCCCGGNNGCTCTNTCTCTATCTTNCTCTTGGCACTTCTCTCGTGCC
TTACAACACCAGCCTCCGCGGCGCATTATACCAACAAGTCTGGCCTGTACCATCTCACCAACGACT
GCCCCAACAGCAGCATCGTTTATGAGGCGGAGACACTGATTTTGCACTTGCCTGGGTGTGTACCTT
GTGTGAAGRTGRACAATCAATCCCGGTGCTGGGTGCAGGCCTCCCCGACCCTGGCAGTGCCGAACG
CGTCTACGCCAGTCACCGGGTTCCGCAAACATGTGGACATCATGGTGGGCGCTGCCGCGTTCTGTT
CAGCTATGTATGTGGGGGACCTGTGCGGGGGCCTTTTCCTCGTTGGACAGCTCTTCACGCTCAGGC
CTCGGATGCATCAGGTTGTCCAGGAGTGTAACTGTTCCATCTACACAGGGCATATCACTGGACACC
GAATGGCA

SEQ ID NO. 47 (VN12, 7d)
ATGAGCACACTTCCAAAACCCCAAAGAAAAACCAAAAGAAACACAAACCGTCGCCCAATGGATGTC
AAGTTCCCGGGCGGCGGTCAGATCGTTGGTGGAGTCTACTTGTTACCGCGCAGGGGCCCACGTTTG
GGTGTGCGCGCGACGAGGAAGACTTCGGAACGGTCCCAGGCCAGAGGTAGGCGCCAACCAATACCC
AAGGTGCGCCAGAACCAAGGCCGAACCTGGGCTCAGCCTGGGTACCCCTGGCCCCTTTATGGGAAC
GAGGGCTGCGGCTGGGCGGGGTGGCTCTTGTCCCCCCGTGGCTCTCGCCCGGACTGGGGNCCCAAT
GACCCCCGGNGGAGGTCCCGCAACCTGGGTAAGGTCATCG
ACACCCTCACTTGCGGCTTCGCCGACCTCATGGAGTACATCCCTGTCGTTGGCGCCCCCCCT
TGGAGGCGTTGCGGCGGAACTGGNACATGGTGTCAGGGCCATCGAGGACGGGATAAACTATGCAAC
AGGGAATCTTCCTGGTTGCTCTTTCTCTATCTTCCWCTTGGCACTTCTCTCGTGCCTCACCACGCC
TGCCTCCGCACTAAACTATGCTAACAAGTCTGGGCTGTATCATCTAACCAATGACTGCCCCAATAG
CAGCATTGTGTATGAGGCGAATGGCATGATCCTGCATCTCCCGGGTTGCGTCCCCTGCGTGAAGAC
CGGCAACCTGACCAAGTGTTGGCTGTCGGCCTCCCCGACATTGGCGGTGCAGAATGCGTCGGTGTC
CATCAGGGGTGTCCGCGAGCACGTGGACCTCTTGGTGGGTGCTGCTGCGTTCTGCTCTGCCATGTA
CGTGGGCGACTTATGCGGTGGGCTCTTTCTCGTTGGGCAGTTGTTCACGTTCAGACCCAGGATGTA
TGAGATCGCCCAGGACTGCAACTGTTCCATCTATGCAGGCCACATCACTGGGCACCGGATGGCG

SEQ ID NO. 41 (FR1, 9a)
ATGAGCACACTTCCAAAACCCCAAAGAAAAACCAAAAGAAATACTAACCGTCGCCCTATGGAC
GTCAAGTTCCCGGGCGGCGGCCAGATCGTTGGTGGAGTTTACTTGTTGCCGCGCAGGGGC
CCTCGTTTGGGTGTGCGCGCGACGAGAAAGACCTCCGAACGGTCCCAGCCTAGAGGCAGG
CGCCAGCCCATACCAAAGGTACGCCAGCCGACAGGCCGTAGCTGGGGTCAACCCGGCTAC
CCTTGGCCCCTTTATGGCAACGAGGGCTGCGGATGGGCGGGATGGCTCCTGTCCCCCCGC
GGGTCTCGTCCTAATTGGGGCCCCAACGACCCCCGGCGAAGGTCCCGCAACTTGGGTAAG
GTCATCGATACCCTTACATNCGGNCTAGCCGACCTCATGGGGTACATCCCTGTCCTAGGAGG
GCCGCTTGGCGGCGTTGCGGCTGCCCTGGCGCATGGCGTTAGGGCAATCGAGGACGGGGTCAATTA
CGCAACAGGGAATCTTCCTGGTTGCTCCTTTTCTATCTTCCTCTTAGCACTGTTATCGTGCCTCAC
TACACCAGCCTCAGCAATTCAAGTCAAGAACGCCTCTGGGATCTACCATCTTACCAATGACTGCTC
GAACAACAGCATCGTTTTTGAGGCGGAGACCATGATACTGCATCTTCCAGGTTGTGTCCCATGTAT
CAAGGCGGGGAATGAGTCACGATGTTGGCTCCCTGTCTCCCCCACCTTAGCCGTCCCCAACTCATC
AGTGCCAATCCACGGGTTTCGCCGACACGTAGACCTCCTCGTTGGGGCAGCGGCATTTTGTTCGGC
CATGTACATCGGAGACCTCTGTGGTAGCATAATCTTGGTAGGGCAGCTTTTTACTTTCAGGCCTAA
GTACCATCAGGTTACCCAGGATTGTAACTGCTCTATNAACNCTGGCCACGTCACGGGACACAGGAT
GGCA

Figure 3 - continued

SEQ ID NO. 49 (NE98, 10a)
ATGAGCACACTTCCTAAACCACAAAGAAAAACCAAAAGAAACACCAACC?CCGGCCACAGGACGTT
AAGTTCCCAGGCGGCGGTCAGATCGTTGGTGGAGTTTACGTGCTACCACGCAGGGGCCCCCAGTTG
GGTGTGCGTGCAGTGCGCAAGACTTCCGAGCGGTCGCAACCTCGCAGTAGGCGCCAACCCATCCCC
AGGGCGCGCCGAACCGAGGGCAGGTCCTGGGCTCAGCCCGGGTACCCTTGGCCCCTATATGGGAAT
GAGGGCTGCGGGTGGGCAGGGTGGCTCCTGTCCCCGCGCGGCTCTC

SEQ ID NO. 51 (NE98, 10a)
GACGGAATTAATTTCGCAACAGGGAATTTACCTGGTTGCTCTTTCTCTATCTTCCTTCTGGCTTTG
TTCTCATGCTTGCTTACACCCACAGCCGGGCTGGAGTACCGTAATGCCTCCGGACTCTACATGGTA
ACTAACGACTGCAGTAACGGTAGTATCGTGTATGAGGCCGGGGATATTATCCTCCACTTACCTGGC
TGTGTCCCCTGCGTACGCTCTGGCAATACATCAAGATGCTGGATCCCTGTGAGCCCYACCGTCGCC
GTGAAGTCGCCCTGCGCCGCCACCGCCTCTCTCCGCACGCACGTGGATATGATGGTGGGRGCGGCC
ACCCTATGCTCAGCTCTCTACGTAGGAGACCTTTGTGGAGCGCTATTTCTTGTYGGGCAGGGGTTC
TCATGGAGACATCGCCAGCATTGGACTGTCCAGGACTGCAACTGTTCCATC

SEQ ID NO. 53 (BNL1,1d)
CTCGACAGTTACTGAGAATGACATCCGTGTCGAGGAATCAATATACCAATGTTGTGACTTGGCCCC
CGAGGCTCGCAAGGCCATAAAGTCGCTCACCGAGCGGCTGTACATCGGGGGCCCYCTAACCAATTC
AAAAGGACAGAACTGCGGCTACCGTCGGTGCCGCGCCAGCGGCGTGCTGACTACCAGCTGCGGCAA
CACCCTGACATGCTACTTGAAAGCCAGAGCGGCCTGTCGAGCTGCAAAGCTCCGGGACTGCACCAT
GCTCGTGTGCGGGGATGACCTTGTCGTTATCTGTGAGAGTGCGGGAGTCGAGGAAGACGCGGCGAA
CCTACGAGCT

SEQ ID NO. 55 (BNL2,1d)
CTCGACAGTTACTGAGAACGACATCCGTACCGAGGRATCAATCTATCAATGTTGTGACTTGGCCCC
YGAGGCCCGCAAGGCCATAAAGTCGCTCACCGAGCGGCTGTACGTCGGGGGCCCCCTAACCAATTC
AAAGGGGCAGAACTGCGGCTATCGTCGGTGTCGCGCTAGCGGCGTGCTGACCACCAGCTGCGGCAA
CACCCTCACATGCTACTTGAAAGCCAGGGCGGCCTGTCGAGCTGCAAAGCTCCAGGACTGCACGAT
GCTCGTGTGCGGAGACGACCTTGTCGTTATCTGTGAGAGCGCGGGAGTCGAGGAGGACGCGGCGAA
CCTACGAGTC

SEQ ID NO. 57 (FR17,1d)
CTCGACAGTTACTGAGAACGACATTCGTGTCGAGGAATCAATCTACCAGTGCTGTGACTTGGCCCC
CGAGGCCCGCAAGGCCATAAAGTCGCTCACCGAGCGGCTGTATATCGGGGGTCCCCTAACCAACTC
AAAAGGGCAGAACTGCGGCTACCGTCGGTGCCGCGCCAGCGGCGTGCTGACTACCAGCTGCGGTAA
TACCCTCACATGTTACTTGAAAGCCAGGGCGGCCTGTCGAGCTGCGAAGCTCCAGGACTGCACAAT
GCTCGTGTGCGGAGACGACCTTGTCGTTATCTGTGAGAGTGCRGGAGTCGAGGAGGATGCGGCGAA
CCTACGAGTC

SEQ ID NO. 59 (CAM1078,1e)

CGTACAGCCTCCAGGACCCCCCCTCCCGGGAGAGCCATAGTGGTCTGCGGAACCGGTGAG
TACACCGGAATTGCCAGGACGACCGGGTCCTTTCTTGGATCAACCCGCTCAATGCCTGGA
GATTTGGGCGTGCCCCCGCAAGACTGCTAGCCGAGTAGTGTTGGGTCGCGAAAGGCCTTG
TGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCACCAT
GAGCACGAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACCAACCGCCGCCCACAGGA
CGTCAAGTTCCCGGGCGGTGGCCAGATCGTTGGTGGAGTCTACGTGCTACCGCGCAGGGG
CCCTAGATTGGGTGTGCGCGCAGCGCGGAAGACTTCGGAGCGGTCGCAACCTCGTGGGAG
GCGCCAACCTATTCCCAAGGAGCGCCGACCCGAGGGCAGGTCCTGGGCGCAGCCCGGGTA
CCCCTGGCCCCTCTATGGTAACGAGGGCTGCGGGTGGGCAGGTNGGCTCCTGTCCCCTCG
CGGCTCCCGTCCTAGTTGGGGTCCTACTGACCCCCGGCGTAGGTCACGCAATTTGGGTAA
GGTCATCGATACCCTCACGTGTTGNTTCGCCGACCTCATGGGGTACATACCG

Figure 3 - continued

SEQ ID NO. 61 (CAM1078,1e)

CTCAACGGTCACTGAAGCTGATATCCGAACAGAGGAGTCCATATACCAATGCTGTGACCTGCACCC
CGAAGCACGTGTAGCCATCAAGTCTTTGACTGAAAGGCTGTACGTCGGGGGGCCCTTGACCAATTC
AAAAGGGGAGAACTGCGGCTATCGCAGATGCCGTGCCAGCGGCGTCTTGACAACCAGCTGCGGCAA
CACCCTCACCTGCTATATCAAGGCCCTAGCAGCCTGTAGAGCTGCCAAGCTCCAGGACTGCACCAT
GCTCGTCTGTGGCGACGACCTGGTCGTGATCTGCGAGAGTGTAGGGACCCAGGAGGATGCGGCGAG
CCTGCGAGCC

SEQ ID NO. 63 (FR2, 1f)

NTCAACAGTCACTGAGAGTGATATCCGTACAGAGGAGTCCATCTACCAATGCTGTGATCTAGACCC
CGAGGCTCGCAAGGCCATAAGGTCCCTCACAGAGAGGCTTTATATCGGGGGTCCCCTGACAAACTC
AAAAGGGCAGAACTGCGGCTACCGCCGATGCCGTGCAAGCGGCGTCCTGACGACTAGCTGCGGCAA
CACCCTCACCTGTTACATAAAGGCCAGGGCAGCCTGTCGAGCTGCGAAGCTCCAGGATTGCTCAAT
GCTCGTCTGTGGCGACGACCTTGTCGTTATCTGCGAGATCGAGGGGNTCCANGAGGATCCGTCGAN
NNNNNNNNNN

SEQ ID NO. 65 (FR16,1g)

CGTAGACCGTGCACCATGAGCACGAATCCTAAACCTCAAAGAAAAACCAAACGTAACATC
AACCGCCGCCCACAGGACGTCAAGTTCCCGGGCGGTGGCCAGATCGTCGGTGGAGTTTAC
CTGTTGCCGCGCAGGGGCCCTAGATTGGGTGTGCGCGCGACTAGGAAGACTTCCGAGCGG
TCGCAACCTCGTGGGAGGCGACAGCCTATCCCCAAGGCTCGCCGATCCGAGGGCAGGTCC
TGGGCTCAGCCCGGGTACCCTTGGCCCCTCTATGGCAATGAGGGCATGGGTTGGGCAGGG
TGGCTCCTGTCCCCCCATGGCTCCCGGCCTAGTTGGGGCCCTTCAGACCCCCGGCGTAGG
TCGCGTAATTTGGGTAAGGTCATCGATACCCTCACATGCGGCTTCGCCGACCTCATGGGG
TACATTCCGCTCGTCGGCGCCCCCCTAGGGGGCGTTGCCAGGGCCCTGGCGCAAGGCTTC
CGGGATCTACCACGTCACCAACGATTGTTCCAATGGGAGCATTGTGTATGAGGCGGAAGG
CATGATCATGCATCTCCCCGGGTGCGTGCCCTGCGTTCGGGAAGGTAATATCTCTCGTTG
CTGGGTACCGTTTTCCCCCACGCTCGCAGCCAGGAATGCTAGCGTCCCCACTCAGGCAAT
TCGGCGACACGTCGACTTGCTTGTTGGGGCGGCCACACTCTGTTCTGCTATGTATGTGGG
GGACCTCTGTGGGTCCGTCTTCCTCGTCGGCCAACTGTTCACCTTCACAWCCGCCAGNA
CTACACAGTGCAAGACTGCAATTGTTCCATCTACCCCGGCCATATAACGGG

SEQ ID NO. 67 (FR16,1g)

NNNNNNNGTCACTGAGAGTGATATCCGTGTCGAGGARTCAATTTACCAATGCTGTGACCTGGCCCC
CGAGGCTCGCGTAGCCATAAAGTCGCTCACTGAGCGGCTATATGTCGGGGGCCCTCTCACCAACTC
AAAAGGACAGAACTGCGGCTATCGCCGGTGCCGTGCGAGCGGTGTGCTGACTACTAGCTGCGGTAA
CACCCTCACATGCTACCTGAAAGCCGCCGCGGCCTGTCGAGCTGCAAAGCTCCGGGAATGCACAAT
GCTCGTGTGTGGCGACGACCTCGTCGTTATCTGTGAGAGTGCGGGGGTCCAGGAGGATGCTGCAAG
CCTNNNNNNN

SEQ ID NO. 69 (BNL3,2e)

CTCGACAGTCACAGAGAGAGATATAAGNACTGAGGAGTCCATATACCAGGCTTGTTCCTTACCCGA
GCAGGCCAGAACTGCCATACACTCATTGACTGAGAGACTCTACGTAGGAGGGCCCATGATGAACAG
CAAAGGGCAATCCTGCGGATACAGGCATTGCCGCGCCAGCGGAGTGCTCACCACCAGTATGGGGAA
TACCATCACGTGCTACATCAAGGCCCTAGCGGCTTGTAAAGCAGCAGGAATAGTGGCCCCCACCAT
GCTGGTGTGCGGCGATGACCTAGTTGTCATCTCAGAGAGTCAGGGAGTCGAGGAGGACGACCGGAA
CCTGANNNNN

Figure 3 - continued

SEQ ID NO. 71 (FR4, 2f)

CTCAACCGTCACAGAGAGGGATATAAGAACTGAGGAGTCCATATACCTGGCCTGCTCCTTACCCGA
GCAGGCCCGGACTGCCATACATTCATTAACTGAGAGACTTTACGTGGGAGGGCCCATGATGAACAG
CAAAGGGCAGTCCTGCGGATACAGGCGTTGCCGCGCTAGCGGAGTGCTCACCACCAGTATGGGGAA
CACCATCACGTGTTATGTGAAAGCCCTCGCAGCTTGTAAAGCTGCGGGCATTGTTGCCCCCACGAT
GCTGGTGTGCGGCGATGACCTGGTTGTCATCTCAGAGAGTCAGGGGGCTGAGGAGGACGAGCGAAA
CCTGAGAGTC

SEQ ID NO. 73 (BNL5,2h)

CTCAACAGTCGCGGAGAGAGACATCAGGACCGAGGAGTCCATTTACCTTGCCTGCTCCTTACCCGA
GCAAGCCCGAACTGCCATACATTCATTGACTGAGAGACTTTACGTAGGAGGGCCCATGATGAACAG
CAAGGGACAGTCCTGCGGTTACAGACGTTGCCGCGCCAGCGGAGTGCTCACCACCAGCATGGGGAA
TACCATCACATGCTATGTGAAGGCATTAGCTGCCTGCAAAGCTGCAGGCATCGTTGCTCCCACGAT
GCTGGTTTGTGGCGACGATCTGGTCATCATCTCAGAGAGTCAGGGAACCGAGGAGGATGAGCGGAA
CCTGAGAGTC

SEQ ID NO. 75 (FR13,2k)

CGNACANCCTCCAGGCCCCCCCCTCCCGGGAGAGCCATAGTGGTCTGCGGAACCGGTGAG
TACACCGGAATTGCCGGGAAGACTGGGTCCTTTCTTGGATAAACCCACTCTATGCCCGGC
CATTTGGGCGTGCCCCCGCAAGACTGCTARCCGAGTAGCGTTGGGTTGCGAAAGGCCTTG
TGGTACTGCCTGATAGGGTGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCATCAT
GAGCACAAATCCTAAACCTCAAAGAAAAACCAAAAGAAACACTAACCGCCGCCCACAGGA
CGTTAAGTTCCCGGGCGGTGGCCAGATCGTTGGCGGAGTATACTTGTTGCCNTGCAGGGG
NCCCAGGTNGNGTNTATGCGCAACGANGAAGACTNCCGAACAGTCCCAGCCACGTGGGAG
GCGCCAGCCCATCCCGAAAGATCGGNGCACCACTGGCAAGTCCTGGGGACGTCCAGGATA
TCCCTGGCCCCTGTATGGGAACGAGGGCCTCGGGTGGGCAGGGTGGCTCCTGTCCCCCCG
GGGCTCCCGCCCGTCATGGGGCCCCACGGACCCCCGGCATAGGTCGCGCAACTTGGGTAA
GGTCATCGATACCCTCACGTNCGGCTTTNCCGACCTCATGGGGTACATTCCCGTCGTTGG
CGCCCCAGTAGGNGGCGTCGCCAGAGCTCTCGCGCATGGCGTGAGAGTCCTGGAGGACGG
GATAAACTATGAAACAGGGAACCTCCCCGGTTGCTCTTTCTCTATCTCCCTCCTTGCTCT
TCTGTCCTGAATTACCGNGCCAGTTTCTGCTGTGGAAATCAAAAACACCAGMAACACATA
CATGGTGACTAACGACTGTTCAAACAGYAGCATCACCTGGCAGCTTNNGNNCGCGGTGCT
TCACGTTCCTGGATGCGTCCCCTGTGAACGAGAGGGCAACAGTTCCCGGTGCTGGATTCC
AGTCACGCCCRACGTAKNCGTGAGCCGACCTGGTGCCCTAACCGAGGGTTTGCGATCGCA
CATCGACACCATCGTAGCGTCCGCAACATTTTGTTCTGCCCTCTACATAGGGGATGTATG
TGGCGCGATAATGATAGCTGCCCAAGTGGTCATCGTCTCGCCGGAGCATCATCACTTTGT
CCAGGACTGTAACTGTTCCATCTACCCGGGCCACATAACGGGGCCTCGTATGTNG

SEQ ID NO. 77 (FR13,2k)

ATCCACAGTCACTGAAAGAGACATCAGAGTTGAAGAGTCCGTTTATCTGTCCTGTTCACTTCCCGA
GGAGGCCCGAGCTGCCATACACTCACTAACTGAGAGGCTGTACGTGGGAGGTCCCATGCAGAACAG
CAAGGGGCAATCCTGCGGATACAGGCGCTGCCGCGCCAGCGGGGTGCTCACCACTAGCATGGGGAA
TACTCTCACATGCTACTTGAAGGCCCAGGCGGCCTGCAGGGCCGCGGGCATTGTTGCACCCACAAT
GCTGGTGTGTGGCGACGACCTGGTCGTCATCTCAGAGAGTCAGGGGACTGAGAGGGACGAGAACAA
CCTGAGACCT

Figure 3 - continued

SEQ ID NO. 79 (FR18,21)

CTCAACAGTCACGGAGAGGGACATCAGGAATGAGGAGTCCATATTCCTGGCCTGCTCGTTGCCCGA
GGAGGCCCGGACTGTCATACATTCGCTCACTGAGAGACTCTACATAGGCGGGCCGATGATGAACAG
CAAAGGCCAGTCCTGTGGATACAGGCGTTGTCGCGCCAGCGGGGTGTTCACCACTAGCATGGGCAA
TACCATCACGTGCTATGTGAAAGCCATGGCAGCTTGCAGAGCTGCCGGGATTGACGCCCCCACAAT
GTTGGTATGTGGCGACGACCTGGTGGTCATCTCAGAGAGTCAGGGGACCGAGGAGGACGAGCGAAA
TCTGAGAGTC

SEQ ID NO. 81 (PAK64,3g)

CTCTTGACTCTACTGTCACTGAACAGGATATCAGGGTAGAAGAAGAAATATACCAATGTTGTGACC
TTGAGCCGGAGGCTAGACGGGCAATCAAATCGCTCACGGAACGGCTTTACGTTGGAGGTCCCATGT
TCAACAGCAAGGGGCTCAAATGCGGATATCGCCGTTGCCGTGCTAGCGGTGTATTGCCCACTAGCT
ACGGTAATACAATCACCTGCTACATCAAGGCCAGAGCGGCTGCTCGAGCTGCGGGCCTTCAAGACC
CATCATTCCTTGTCTGCGGAGATGATTTGGTGGTAGTGGCTGAGAGTTGCGKCGTTGATGAGGAGG
ATAGGGCAGC

SEQ ID NO. 83 (BNL8,4k)

CTCCACTGTAACCGAAAAGGACATCAGGCCCGAGGAAGAGGTCTATCAGTGTTGTGACCTGGAGCC
CGAAGCTCGCAAGGTTATTACCGCCCTCACAGAAAGACTCTACGTGGGCGGCCCCATGCACAACAG
CAAGGGAGACCTTTGTGGGTATCGGAGATGCCGCGCAAGCGGCGTCTACACGACCAGCTTCGGAAA
CACACTGACGTGCTACCTCAAAGCCTCAGCTGCTATTAGAGCGGCAGGGCTGAGAGACTGCACCAT
GCTGGTTTGCGGTGACGACTTGGTCGTCATCGCTGAGAGCGATGGCGTAGAGGAGGATAACCGAGC
CCTCCNAGCC

SEQ ID NO. 85 (BNL12,41)

CTCCACGGTGACTGAAAAGGACATCAGGGTCGAGGAAGAGATCTATCAATGTTGTGACCTGGARCC
CGAAGCCCGCAAAGCAATATCCGCCCTCACAGAGAGRCTCTACTTGGGCGGCCCCATGTATAACAG
CAAAGGGGAGCTCTGCGGGTATCGGAGGTGCCGCGCGAGCGGAGTGTACACCACAAGTTTCGGGAA
CACAGTGACCTGCTATCTTAAGGCCACCGCAGCTACCAGGGCTGCAGGCCTAAAAGACTGCACCAT
GCTGGTCTGCGGTGACGACTTGGTCGTCATCGCCGAGAGCGAGGGCGTAGAGGAGGATTCCCAACC
CCTCCGAGCC

SEQ ID NO. 87 (EG81,4m)

CTCCACCGTAACCGAAAGGGACATCAGGGTCGAGGAGGAGGTCTATCAGTGTTGTGATCTGGAGCC
AGAGGCCCGCAAGGCAATATCCGCCCTCACGGAGAGACTCTATGTGGGCGGTCCCATGTTTAACAG
CAAGGGAGACCTATGTGGCTACCGCAGGTGCCGCGCAAGCGGCGTCTACACCACCAGCTTCGGAAA
CACACTGACCTGCTACCTCAAGGCCACGGCCGCTACCAGAGCGGCCGGCCTGAAGGATTGCACAAT
GCTGGTTTGCGGGGACGACCTGGTCGTCATCGCAGAGAGCGATGGCGTGGACGAGGACCGCCGAGC
CCTCCAAGCT

SEQ ID NO. 89 (VN13,7a)

CTCAACAGTCACAGAGCGCGATGTCCAGACGGAGCATGACATCTACCAGTGCTGTAAGTTGGAGCC
CGCAGCACGGACAGCCATCACATCGCTTACTGACCGATTGTACTNCGGTGGTCCCATGTNTAACTC
TAAAGGTCAGGCATGTGGATACCGTAGGTGCAGGGCCAGTGGCGTCTTGACCACCATCCTGGCCAA
TACTCTGACTTGCTACTTGAAAGCTCAGGCGGCATGCAGAGCTGCCGGGCTGAAGGACTTTGACAT
GTTGGTCTGCGGAGACGACCTTGTCGTTATTTCGGAGAGTTTGGGGGTCTCGGAGGACACTAGTGC
ACTGCGAGCT

Figure 3 - continued

SEQ ID NO. 91 (VN4,7c)

CTCGACAGTCACCGAGCGCGACATCCRCACCGAGCACGACATCTACCAATGCTGCCAACTTGACCC
GGTGGCACGCAAGGCTATTACATCTCTGACTGAGCGGCTGTACTGCGGWGGGCCCATGATGAACTC
CCGTGGTCAATCATGTGGATACCGTAGGTGCCGAGCCAGTGGCGTGCTCACCACGAGCTTGGGCAA
TACCCTAACATGCTATTTGAAAGCACAAGCAGCGTGTAGGGCAGCAAAGCTCAAAAACTATGACAT
GTTAGTCTGCGGAGACGATCTAGTCGTTATCGCGGAGAGTGGAGGAGTCTCTGAGGATGTTGACGC
CCTGCGAGCA

SEQ ID NO. 93 (VN12,7d)

CTCCTCCGTCACGGAGCGTGACATCCGCACTGAACACGACATCTATCAGTGCTGCCAATTAGATCC
GGTAGCACGGAAAGCCATTACATCTCTTACTGAGCGGCTGTACTGCGGCGGCCCCATGTACAACTC
TCGAGGTCAGTCATGTGGGTACCGCAGGTGCCGGGCTAGTGGTGTCTTCACCACAAGCTTGGGCAA
CACCATGACATGCTACCTGAAGGCTCAGGCGGCTTGTAGGGCAGCRAAGCTCAAAAACTTTGACAT
GTTGGTCTGCGGAGACGACCTAGTCGTTATTGCTGAGAGCGGAGGAGTCCCTGAGGATGCCGGGGC
CCTGCGAGTC

SEQ ID NO. 95 (FR1,9a)

ATCCACAGTCACGGGGCGCGACATACGCACAGAACNAGACATTTACCTGTCCTGCCAGCTCGACCC
AGAGGCCCGGAAAGCCATAAAGTCTCTCACTGAGAGGCTCTATGTCGGGGGCCCTATGTACAACTC
AAAGGGCCAACTCTGTGGTCAACGCCGATGCCGAGCAAGCGGAGTACTCCCCACAAGCATGGGTAA
CACCATCACATGCTTCCTGAAGGCAACCGCCGCTTGCCGAGCAGCCGGCTTTACAGATTATGACAT
GTTGGTCTGCGGAGACGATTTGGTTGTCGTAACTGAGAGTGCTGGAGTCAACGAGGATATCGCTAA
CCTGCGAGCC

SEQ ID NO. 97 (NE98,10a)

CTCCACTGTCACTGAGCAGGACATCAGGGTAGAACTTTCCATCTTTCAGGCCTGTGACCTCAAGGA
CGAGGCTAGGAGGGTGATAACTTCACTCACGGAGCGGCTTTACTGTGGTGGTCCTATGTTCAACAG
CAAGGGACAACACTGCGGTTACCGCCGCTGCCGTGCTAGTGGGGTGCTACCCACCAGCTTCGGGAA
CACAATCACCTGTTACATCAAAGCAAAGGCAGCTACCAAAGCTGCCGGAATTAAAAATCCATCATT
CCTTGTCTGCGGAGATGACTTGGTCGTGATTGCTGAGAGTGCAGGGATCGATGAGGACAAGAGCGC
CTTGAGAGCT

SEQ ID NO. 99 (FR14,11a)

CTCTACCGTCACAGAGAGGGACATACGGACAGAAGAATCCATCTATCTGTCTTGTCAATTGCCTGA
AGAGGCCCGGAAAGCCATTAAATCGCTGACAGAGAGACTATACGTGGGCGGCCCGATGGAAAACAG
CAAGGGCCAGGCTTGCGGATATAGGCGTTGCCGCGCAAGCGGGGTATTCACCACAAGCTTGGGGAA
CACCATGACTTGTTACATCAAAGCTAAAGCGGCTTGTAAAGCCGCTGGCATTGTAGACCCGGTGAT
GCTCGTGTGCGGTGACGACCTAGTGGTCATCTCAGAAAGCAAGGGGGTGGAGGAGGACCAGCGGGA
CCTACGAGTC

SEQ ID NO. 101 (FR15,11a)

CTCCACTGTCACTGAGAGAGACATACGGACAGAAGAATCCATCTAYYTGGCTTGTCAATTGCCCGA
AGAGGCCCGGAAGGCCATTAAATCACTGACAGAGAGACTATACGTGGGCGGCCCGATGGAAAACAG
CAAAGGCCAGGCCTGCGGATATAGGCGTTGCCGCGCAAGCGGGGTATTCACCACAAGCTTGGGGAA
CACCATGACTTGTTACATCAAGGCCAARGCAGCTTGTAAAGCYGCTGGCATTGTTGACCCGGTGAT
GCTCGTGTGCGGCGACGACCTAGTGGTCATCTCAGAGAGCAAGGGGGTAGAGGAGGACCAGCGAGA
CCTAC

Figure 3 - continued

SEQ ID NO. 103 (FR19,11a)
CGTACAGCCTCCAGGACCCCCCCTCCCGGGAGAGCCATAGTGGTCTGCGGAACCGGTGAGTACACC
GGAATTGCCGGGAAGACTGGGTCCTTTCTTGGATTAACCCACTCTATGCCCGGAGATTTGGGCGTG
CCCCCGCAAGACTGCTAGCCGAGTAGCGTTGGGTTGCGAAAGGCCTTGTGGTACTGCCTGATAGGG
TGCTTGCGAGTGCCCCGGGAGGTCTCGTAGACCGTGCACCATGAGCACGAATCCTAAACCTCAAAG
ACAAACCAAAAGAAACACCAACCGCCGCCCACAGGACGTTAAGTTCCCGGGCGGTGGCCAGATCGT
TGGCGGGGTGTACTTGTTGCCGCGCAGGGGCCCCAGAGTGGGTGTGCGCGCGACGAGAAAGACCTC
GGAGCGGTCCCAGCCGCGTGGGAGGCGCCAACCTATCCCCAAGGTTAGGCGCACCACCGGCCGTT

SEQ ID NO. 105 (FR19,11a)
CTCTACTGTCACAGAGAGGGATATACGAACAGAGGAATCCATYTATCTGGCTTGTCAATTGCCCGA
AGAGGCCCGGAAGGCCATCAAATCACTGACAGAGAGACTATACGTGGGCGGCCCGATGGAAAACAG
CAAGGGCCAGGCCTGCGGATACAGGCGTTGCCGCGCAAGCGGGGTATTCACCACAAGCTTGGGGAA
CACCATGACTTGTTACATCAAAGCCAAGGCGGCTTGTAAAGCCGCTGGCATTGTTGACCCAGTGAT
GCTCGTGTGCGGCGACGACCTAGTGGTCATCTCAGAAAGCAAGGGGGTGGAGGAGGACCAACGAGA
CCTACGANTC

SEQ ID NO. 2 (BNL1, 1d)
MSTNPKPQRKTKRNTNRRPXXXXXPGGGQIVGGVYLLPRRGPRXGVRATRKTSERSQPRGRRQPIP
KAXRXEGRSWAQPGYPWPLYGNEGCGWAXWLLSPRGSRPNWGP

SEQ ID NO. 4 (BNL1, 1d)
DGVNYATGNLPGCSFSIFLLALLSCLTVPXTAHEVRNASGVYHVTNDCSNSSIIYEMDGMIMHYPG
CVPCVREDNHLRCWMALTPTLAVKXASVPTXAIRRHVDLLVGXXTFCSAMYVXDLCGSVFLAGQLF
TFSPRMHHTTQECNCSI

SEQ ID NO. 6 (BNL2, 1d)
MSTNPKPQRKTKRNTNRRPQDVKXPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRDRRQPIP
KARQSDGXXWAQPGHPWPLYGNEGCGWAGWLLSPRGSRPSWGP

SEQ ID NO. 8 (BNL2, 1d)
DGVNYATGNLPGCSFSIFLLAFLSCLTVPTTAHEVRNASGVYHLTNDCSNSSIIYEMSGMILHAPG
CVPCVRENNSSRCWMXLTPTLAVKDANVPTAAIRRHVDLLVGTAAFRSAMYVGDLCGSVFLVGQLF
TFSPRLYHTTQECNCSI

SEQ ID NO. 10 (CAM1078, 1e)
MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYVLPRRGPRLGVRAARKTSERSQPRGRRQPIP
KERRPEGR

SEQ ID NO. 12 (FR2, 1f)
MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGRRQPIP
KARRPEGRSWAQPGYPWPLYANEGCGWAGWLLSPRGSRPSWGPNDPRRRSRNLGKVIDTLTCGFAD
LMGYIPLVGAPLGGASRTLXHGVRVLXGGVXXXXXNLXGCSXXIFLLXLLSCLTVPTSAYEVHSTT
DGYHVTNDCSNGSIVYEAKDIILHTPGXVPCIREGNISRCWVPLTPTLAARIANAPIDEVRRHVDL
LVGAAVFCSAMYIGDLCGGVFLVGQLFTFTSRRHWT
VQDCNCSIYSGHITGHXXX

SEQ ID NO. 14 (BNL3, 2e)
MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGRRQPIP
KDRXATGRSWGRPGYPWPLYGNEGLGWAGWLLSPRGSRPSWG

SEQ ID NO. 16 (BNL3, 2e)
TCXXADLMGYXPVVGAPVGGXARALAXGVRVLEDGINYXTGNLPGCSFSIFXLALLSCVTVPVSXV
EVKNTSQAYMATNDCSNNSIVWQLXDAVLHVPGCVPCENSSGRFHCWIPISPNIAVSKPGALTKGL
RARIDAVVMSATLCSALYVGDVCGAVMIAAQAFIVAPKRHYFVQECNCSIYPGHITGHRMA

Figure 3 - continued

SEQ ID NO. 18 (FR4, 2f)
MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRAPRKTSERSQPRGRRQPIP
KDRRATGKSWGRPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPNDPRHRSRNLGKVIDTLTCGFXD
LMGYIPVVGAPVGGVARALAHGVRVLEDGINYATGNLPGCSFSIFLLALLSCITVPVSAIQVKNNS
HFYMATNDCANDSIVWQLRDAVLHVPGCVPCERSGNRTFCWTAVSPNVAVSRPGALTRGLRAHIDT
IVMSATLCSALYIGDLCGAVMIAAQVAVVSPQYHTFVQECNCSIYPGHITGHRMX

SEQ ID NO. 20 (BNL4, 2g)
DGVNYATGNLPGCSFSIFLLALLSCVTVPVSAVQVKNTSTMYMATNDCSNNSIIWQMQGAVLHVPG
CVPCELQGNKSRCWIPVTPNVAVNQPGALTRGLRTHIDTIVMVATLCSALYIGDVCGAVMIAAQVV
IVSPQHHNFSQDCNCSI

SEQ ID NO. 22 (BNL5, 2h)
MSTNPKPQRKTKRNTNRRPQDVKFPGGGRSLAEYTCARRGKLRRSSMG

SEQ ID NO. 24 (BNL5, 2h)
DGINYATGNLPGCSFSIFLLALLSCLTVPASAVQVKNTSHSYMVTNDCSNSSIVWQLKDAVLHVPG
CVPCERHQNQSRCWIPVTPNVAVSQPGALTRGLRTHIDTIVASATVCSALYVGDFCGAVMLVSQFF
MISPQHHIFVQDCNCSI

SEQ ID NO. 26 (BNL6, 2i)
DGINYATGNLPGCSFSIFLLALLSCITVPVSAVQVANRSGSYMVTNDCSNSSIVWQLEEAVLHVPG
CVPCEWKDNTSRCWIPVTPNIAVSQPGAXTKGLRTHIDIIVASATFCSALYV

SEQ ID NO. 28 (BNL7, 4k)
MSTNPKPQRKTKRNTNRRPMDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGRRQPIP
KARRSEGRSWAQPGYPWPLYGNEGCGWAXWLLSPRGSRPSWGPNDPRRRSR

SEQ ID NO. 30 (BNL7, 4k)
DGINFATGNLPGCSFSIFLLALLSCLTVPASAINYRNVSGIYYVTNDCPNSSIVYEADHHILHLPG
CVPCVREGNQSRCWVALTPTVAAPYIGAPLESLRSHVDLMVGAATVCSALYIGDXCXGLFLVGQMF
SFRPRRHWTTQDCNCSI

SEQ ID NO. 32 (BNL8, 4k)
DGINYATGNLPGCSFSIFLLALLSCLTVPASAINYRNTSGIYHVTNDCPNSSIVYEADHHILHLPG
CVPCVRTGNQSRCWVALTPTVAAPYIGAPLESLRSHVDLMVGAATVCSALYIGDLCGGLFLVGQMF
SFRPRRHWTAQDCNCSI

SEQ ID NO. 34 (BNL9, 4k)
DGINYATGNLPGCSFSIFLLALLSCLTVPASAINYHNTSGIYHITNDCPNSSIVYEADHHILHLPG
CVPCVRVGNQSSCWVALTPTIAAPYIGAPLESLRSHVDLMVGAATVCSALYIGDLCGGAFLVGQMF
SFRPRRHWTTQDCNCSI

SEQ ID NO. 36 (BNL10, 4k)
DGINYATGNIPGCXFSIFLXALLSCLTVPASATNYRNVSGIYHVTNDCPNSSIVYEADHHILALPG
CVPCVRVGNQSRCWVALTPTVAAPYTAAPLESLRSHVDLMVGAATVCSALYIGXLCGGLFLVGQMF
SXQPRRHWTTQDCNCSI

SEQ ID NO. 38 (BNL11, 4k)
DGINYATGXLPGCSFSIFLLALLSCLTVPASATNYRNVSGIYHVTNDCPNSSIVFEADHHILHLPG
CVPCVKEGNHSRCWVALTPTVAAPYIGAPLESLRSHVDVMVGAATVCSALYIGDLCGGLFLVGQMF
SFRPRRHWTTQECNCSI

SEQ ID NO. 40 (BNL12, 41)
DGINYATGNLPGCSFSIFILALLSCLTVPASAQHYRNVSGIYHVTNDCPNSSIVYESDHHILHLPG
CVPCVKTGNTSRCWVALTPTVAAPILSAPLMSVRRHVDLMVGAATLSSALYVGDLCGGAFLVGQMF
TFQPRRHWTVQDCNCSI

Figure 3 - continued

SEQ ID NO. 46 (VN13, 7a)

MSTLPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGRRQPIP
KVRHQTGRTWAQPGYPWPLYGNEGCGWAGWLLSPXGSRPNWGPNDPRXRSRNLGKVIDTLTXXFAD
LIEYI

SEQ ID NO. 44 (VN4, 7c)

MSTLPKPQRKTKRNTIRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGRRQPIP
KVRHQTGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPNWGPNDPRRRSRNLGKVIDTLTCGFAD
LMGYIPVVGAPXGGVAXALAHGVXXIEDXVNYATXNLPXXSXSIXLLALLSCLTTPASAAHYTNKS
GLYHLTNDCPNSSIVYEAETLILHLPGCVPCVKXXNQSRCWVQASPTLAVPNASTPVTGFRKHVDI
MVGAAAFCSAMYVGDLCGGLFLVGQLFTLRPRMHQVVQECNCSIYTGHITGHRMA

SEQ ID NO. 48 (VN12, 7d)

MSTLPKPQRKTKRNTNRRPMDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQARGRRQPIP
KVRQNQGRTWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPDWXPNDPRXRSRNLGKVIDTLTCGFAD
LMEYIPVVGAPLGGVAAELXHGVRAIEDGINYATGNLPGCSFSIFXLALLSCLTTPASALNYANKS
GLYHLTNDCPNSSIVYEANGMILHLPGCVPCVKTGNLTKCWLSASPTLAVQNASVSIRGVREHVDL
LVGAAAFCSAMYVGDLCGGLFLVGQLFTFRPRMYEIAQDCNCSIYAGHITGHRMA

SEQ ID NO. 42 (FR1, 9a)

MSTLPKPQRKTKRNTNRRPMDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGRRQPIP
KVRQPTGRSWGQPGYPWPLYGNEGCGWAGWLLSPRGSRPNWGPNDPRRRSRNLGKVIDTLTXXLAD
LMGYIPVLGGPLGGVAAALAHGVRAIEDGVNYATGNLPGCSFSIFLLALLSCLTTPASAIQVKNAS
GIYHLTNDCSNNSIVFEAETMILHLPGCVPCIKAGNESRCWLPVSPTLAVPNSSVPIHGFRRHVDL
LVGAAAFCSAMYIGDLCGSIILVGQLFTFRPKYHQVTQDCNCSXNXGHVTGHRMA

SEQ ID NO. 50 (NE98, 10a)

MSTLPKPQRKTKRNTNXRPQDVKFPGGGQIVGGVYVLPRRGPQLGVRAVRKTSERSQPRSRRQPIP
RARRTEGRSWAQPGYPWPLYGNEGCGWAGWLLSPRGSRPSWGPNDPRRR

SEQ ID NO. 52 (NE98, 10a)

DGINFATGNLPGCSFSIFLLALFSCLLTPTAGLEYRNASGLYMVTNDCSNGSIVYEAGDIILHLPG
CVPCVRSGNTSRCWIPVSXTVAVKSPCAATASLRTHVDMMVXAATLCSALYVGDLCGALFLXGQGF
SWRHRQHWTVQDCNCSI

SEQ ID NO. 54 (BNL1,1d)

STVTENDIRVEESIYQCCDLAPEARKAIKSLTERLYIGGXLTNSKGQNCGYRRCRASGVLTTSCGN
TLTCYLKARAACRAAKLRDCTMLVCGDDLVVICESAGVEEDAANLRA

SEQ ID NO. 56 (BNL2,1d)

STVTENDIRTEXSIYQCCDLAXEARKAIKSLTERLYVGGPLTNSKGQNCGYRRCRASGVLTTSCGN
TLTCYLKARAACRAAKLQDCTMLVCGDDLVVICESAGVEEDAANLRV

SEQ ID NO. 58 (FR17,1d)

STVTENDIRVEESIYQCCDLAPEARKAIKSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGN
TLTCYLKARAACRAAKLQDCTMLVCGDDLVVICESXGVEEDAANLRV

Figure 3 - continued

SEQ ID NO. 60 (CAM1078,1e)

MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYVLPRRGPRLGVRAARKTSERSQPRGRRQPIP
KERRPEGRSWAQPGYPWPLYGNEGCGWAGXLLSPRGSRPSWGPTDPRRRSRNLGKVIDTLTCXFAD
LMGYIP

SEQ ID NO. 62 (CAM1078,1e)

STVTEADIRTEESIYQCCDLHPEARVAIKSLTERLYVGGPLTNSKGENCGYRRCRASGVLTTSCGN
TLTCYIKALAACRAAKLQDCTMLVCGDDLVVICESVGTQEDAASLRA

SEQ ID NO. 64 (FR2, 1f)

STVTESDIRTEESIYQCCDLDPEARKAIRSLTERLYIGGPLTNSKGQNCGYRRCRASGVLTTSCGN
TLTCYIKARAACRAAKLQDCSMLVCGDDLVVICEIEGXXEDPSXXXX

SEQ ID NO. 66 (FR16,1g)

MSTNPKPQRKTKRNINRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATRKTSERSQPRGRRQPIP
KARRSEGRSWAQPGYPWPLYGNEGMGWAGWLLSPHGSRPSWGPSDPRRRSRNLGKVIDTLTCGFAD
LMGYIPLVGAPLGGVARALAQGFRDL

SEQ ID NO. 68 (FR16,1g)

XXVTESDIRVEXSIYQCCDLAPEARVAIKSLTERLYVGGPLTNSKGQNCGYRRCRASGVLTTSCGN
TLTCYLKAAAACRAAKLRECTMLVCGDDLVVICESAGVQEDAASXXX

SEQ ID NO. 70 (BNL3,2e)

STVTERDIXTEESIYQACSLPEQARTAIHSLTERLYVGGPMMNSKGQSCGYRHCRASGVLTTSMGN
TITCYIKALAACKAAGIVAPTMLVCGDDLVVISESQGVEEDDRNLXX

SEQ ID NO. 72 (FR4, 2f)

STVTERDIRTEESIYLACSLPEQARTAIHSLTERLYVGGPMMNSKGQSCGYRRCRASGVLTTSMGN
TITCYVKALAACKAAGIVAPTMLVCGDDLVVISESQGAEEDERNLRV

SEQ ID NO. 74 (BNL5,2h)

STVAERDIRTEESIYLACSLPEQARTAIHSLTERLYVGGPMMNSKGQSCGYRRCRASGVLTTSMGN
TITCYVKALAACKAAGIVAPTMLVCGDDLVIISESQGTEEDERNLRV

SEQ ID NO. 76 (FR13,2k)

MSTNPKPQRKTKRNTNRRPQDVKFPGGGQIVGGVYLLXCRXPRXXXCATXKTXEQSQPRGRRQPIP
KDRXTTGKSWGRPGYPWPLYGNEGLGWAGWLLSPRGSRPSWGPTDPRHRSRNLGKVIDTLTXGFXD
LMGYIPVVGAPVXGVARALAHGVRVLEDGINYETGNLPGCSFSISLLALLSITXPVSAVEIKNTXN
TYMVTNDCSNXSITWQLXXAVLHVPGCVPCEREGNSSRCWIPVTPXVXVSRPGALTEGLRSHIDTI
VASATFCSALYIGDVCGAIMIAAQVVIVSPEHHHFVQDCNCSIYPGHITGPRMX

SEQ ID NO. 78 (FR13,2k)

STVTERDIRVEESVYLSCSLPEEARAAIHSLTERLYVGGPMQNSKGQSCGYRRCRASGVLTTSMGN
TLTCYLKAQAACRAAGIVAPTMLVCGDDLVVISESQGTERDENNLRP

Figure 3 - continued

SEQ ID NO. 80 (FR18,21)

STVTERDIRNEESIFLACSLPEEARTVIHSLTERLYIGGPMMNSKGQSCGYRRCRASGVFTTSMGN
TITCYVKAMAACRAAGIDAPTMLVCGDDLVVISESQGTEEDERNLRV

SEQ ID NO. 82 (PAK64,3g)

STVTEQDIRVEEEIYQCCDLEPEARRAIKSLTERLYVGGPMFNSKGLKCGYRRCRASGVLPTSYGN
TITCYIKARAAARAAGLQDPSFLVCGDDLVVVAESCXVDEEDRAALR

SEQ ID NO. 84 (BNL8,4k)

STVTEKDIRPEEEVYQCCDLEPEARKVITALTERLYVGGPMHNSKGDLCGYRRCRASGVYTTSFGN
TLTCYLKASAAIRAAGLRDCTMLVCGDDLVVIAESDGVEEDNRALXA

SEQ ID NO. 86 (BNL12,41)

STVTEKDIRVEEEIYQCCDLXPEARKAISALTEXLYLGGPMYNSKGELCGYRRCRASGVYTTSFGN
TVTCYLKATAATRAAGLKDCTMLVCGDDLVVIAESEGVEEDSQPLRA

SEQ ID NO. 88 (EG81,4m)

STVTERDIRVEEEVYQCCDLEPEARKAISALTERLYVGGPMFNSKGDLCGYRRCRASGVYTTSFGN
TLTCYLKATAATRAAGLKDCTMLVCGDDLVVIAESDGVDEDRRALQA

SEQ ID NO. 90 (VN13,7a)

STVTERDVQTEHDIYQCCKLEPAARTAITSLTDRLYXGGPMXNSKGQACGYRRCRASGVLTTILAN
TLTCYLKAQAACRAAGLKDFDMLVCGDDLVVISESLGVSEDTSALRA

SEQ ID NO. 92 (VN4,7c)

STVTERDIXTEHDIYQCCQLDPVARKAITSLTERLYCXGPMMNSRGQSCGYRRCRASGVLTTSLGN
TLTCYLKAQAACRAAKLKNYDMLVCGDDLVVIAESGGVSEDVDALRA

SEQ ID NO. 94 (VN12,7d)

SSVTERDIRTEHDIYQCCQLDPVARKAITSLTERLYCGGPMYNSRGQSCGYRRCRASGVFTTSLGN
TMTCYLKAQAACRAXKLKNFDMLVCGDDLVVIAESGGVPEDAGALRV

SEQ ID NO. 96 (FR1,9a)

STVTGRDIRTEXDIYLSCQLDPEARKAIKSLTERLYVGGPMYNSKGQLCGQRRCRASGVLPTSMGN
TITCFLKATAACRAAGFTDYDMLVCGDDLVVVTESAGVNEDIANLRA

SEQ ID NO. 98 (NE98,10a)

STVTEQDIRVELSIFQACDLKDEARRVITSLTERLYCGGPMFNSKGQHCGYRRCRASGVLPTSFGN
TITCYIKAKAATKAAGIKNPSFLVCGDDLVVIAESAGIDEDKSALRA

SEQ ID NO. 100 (FR14,11a)

STVTERDIRTEESIYLSCQLPEEARKAIKSLTERLYVGGPMENSKGQACGYRRCRASGVFTTSLGN
TMTCYIKAKAACKAAGIVDPVMLVCGDDLVVISESKGVEEDQRDLRV

Figure 3 - continued

SEQ ID NO. 102 (FR15,11a)

STVTERDIRTEESIXXACQLPEEARKAIKSLTERLYVGGPMENSKGQACGYRRCRASGVFTTSLGN
TMTCYIKAXAACKXAGIVDPVMLVCGDDLVVISESKGVEEDQRDLXX

SEQ ID NO. 104 (FR19,11a)

MSTNPKPQRQTKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRVGVRATRKTSERSQPRGRRQPIP
KVRRTTGR

SEQ ID NO. 106 (FR19,11a)

STVTERDIRTEESXYLACQLPEEARKAIKSLTERLYVGGPMENSKGQACGYRRCRASGVFTTSLGN
TMTCYIKAKAACKAAGIVDPVMLVCGDDLVVISESKGVEEDQRDLRX

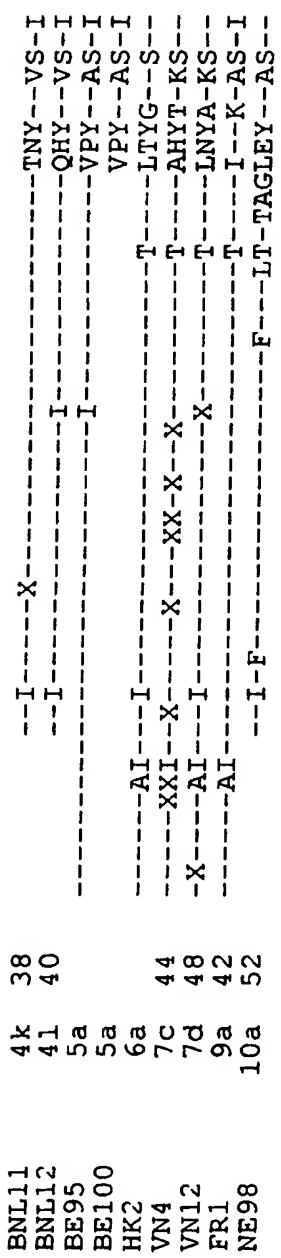
Figure 4. Core/E1 amino acid alignment

Isolate	Type	SEQ ID	1 MSTNPKPQKKNKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATR 50
HCV-1	1a		MSTNPKPQKKNKRNTNRRPQDVKFPGGGQIVGGVYLLPRRGPRLGVRATR
HCV-J	1b		--------R-T----------
BNL1	1d	2	--------R-T--------XXXXX----------X------
BNL2	1d	6	--------R-T----------X----------
CAM1078	1e	10/60	--------R-T----------V----------A-
FR2	1f	12	--------R-T---I----------
FR16	1g	66	--------R-T----------
HC-J6	2a		--------R-T----------
HC-J8	2b		--------R-T----------
CH610	2c		--------R-T----------
NE92	2d		--------R-T----------
BNL3	2e	14	--------R-T----------
FR4	2f	18	--------R-T----------P-
FR13	2k	76	--R-T----I----------V----------C---
EB1	3a		---L----R-T----I----------V----------
NZL1	3a		---L----RQT----L----N----------V----------V-
HCV-TR	3b		--------R-T--------M----------
GB358	4c		--------R-T--------M----------
DK13	4d		--------R-T--------M----------
CAM600	4e		------L-R-T--------M----------
GB809	4e		-----T----------G
HPCCOREEZA	4?		-T--------M----------
HPCCOREZB	4?		-----M----------
HPCCOREZC	4?		--------R-T--------M----------
GB724	4?		--------R-T--------M----------
BNL7	4k	28	--------R-T----------
BE95	5a		---L----R-T--------T----------
HK2	6a		---L----R-T----------
VN13	7a	46	---L----R-T----I----------
VN4	7c	44	---L----R-T--------M----------
VN12	7d	48	---L----R-T--------M----------
FR1	9a	42	---L----R-T-----X----------V------Q-----V-
NE98	10a	50	--------RQT----------V------
FR19	11a	104	

Isolate	Type	SEQ ID	51 core-V 100
			KTSERSQPRGRRQPIPKARRPEGRTWAQPGYPWPLYGNEGCGWAGWLLSP
HCV-1	1a		
HCV-J	1b		----------M---------
BNL1	1d	2	----------X-X---S----------X-----
BNL2	1d	6	---------D---------QSD-XX-----H----------
CAM1078	1e	10/60	----------E------S----------X----
FR2	1f	12	----------S----------A----------
FR16	1g	66	----------S---S----------M---------
HCJ6	2a		----------D--ST-KS-GK----------L---------
HCJ8	2b		----------D--ST-KS-GK----------
CH610	2c		----------D--TT-KS-GR----------L---------
NE92	2d		----------D---T-KS-GK----------L---------
BNL3	2e	14	----------D-XAT--S-GR----------L---------
FR4	2f	18	----------D--AT-KS-GR----------L---------
FR13	2k	76	--X-Q----------D-XTT-KS-GR----------L---------
EB1	3a		----------S---S----------
NZL1	3a		----------S---S----------
HCV-TR	3b		---------KQ-HL-----SR---S----------K---L-------
GB358	4c		----------S---S----------
DK13	4d		----------QL---S----------
CAM600	4e		----------T---S----------
GB809	4e		----------S---S----------
BNL7	4k	28	----------S---S----------X-----
HPCCOREEZA	4?		----------S---S-----F-
HPCCOREZB	4?		----------S---S----------
HPCCOREZC	4?		----------S---S----------K
GB724	4?		----------S---S---A----------
BE95	5a		----------Q-T--S-G---------A---L---------
HK2	6a		----------Q-Q--H----------
VN13	7a	46	----------V-HQT----------
VN4	7c	44	----------V-HQT----------
VN12	7d	48	-------A---------V-QNQ----------
FR1	9a	42	----------V-Q-T--S-G----------
NE98	10a	50	---------S------R---T---S----------
FR19	11a	104	----------V--TT--

Isolate	Type	SEQ ID	101 RGSRPSWGPTDPRRRSRNLGKVIDTLTCGFADLMGYIPLVGAPLGGAARA 150
HCV1	1a		----------
HCV-J	1b		----------
BNL1	1d	2	----N---
BNL2	1d	6	----------
CAM1078	1e	10/60	----------X----------
FR2	1f	12	----------N----------
FR16	1g	66	H---------S----------
HC-J6	2a		----------N---H----V----------
HC-J8	2b		----T---------H-----R----I----------
CH610	2c		----------H----------
NE92	2d		----------H----------
BNL3	2e	14/16	----------XX-----X-V---V--X---
FR4	2f	18	----------N---H----------X----------
FR13	2k	76	----------H----------X--X----------
HCV-TR	3b		----------N----------F----------
GB116	4c		----------
DK13	4d		----------N----------
CAM600	4e		-X--X---N---X----------
GB809	4e		----------N----------
G22	4f		----------
GB549	4g		----------
GB438	4h		----------
BNL7	4k	28	----------N----------
BE95	5a		----N---N----K----------
HK2	6a		----H---N----------
VN13	7a	46	X----N---N---X----------XX----IE--
VN4	7c	44	----N---N----------
VN12	7d	48	----D-X-N---X----------
FR1	9a	42	----N---N----------
NE98	10a	50	----------N-----

Isolate	Type	SEQ ID	151 V1 200
			LAHGVRVLEDGVNYATGNLPGCSFSIFLLALLSCLTVPASAYQVRNSTGL
HCV1	1a		
HCV-J	1b		----------I-----E---VS-I
BNL1	1d	4	----------XT-HE---AS-V
BNL2	1d	8	----------F-------TT-HE---AS-V
FR2	1f	12	-X------XG--XXXXX--X---XX----X--------T---E-HST-DG
FR16	1g	66	--Q-F-D-
HC-J6	2a		----------F----------I-T-V--AE-K-ISTG
HC-J8	2b		----------I----------V---V--VE---ISSS
CH610	2c		----------I----------S-----IS--V--VE-K-TSTS
S83	2c		VE-KDTGDS
NE92	2d		----------I----------I---V-GL--K-TSSS
BNL3	2e	16	--X--------I--X----------X------V---V-XVE-K-TSQA
FR4	2f	18	----------I----------I---V--I--K-NSHF
BNL4	2g	20	----------V---V--V--K-TSTM
BNL5	2h	24	--I----------V--K-TSHS
BNL6	2i	26	--I----------I---V--V--A-RS-S
FR13	2k	76	----------I--E----------S------/I-X-V--VEIK-TXNT
BR36	3a		LEW--TS--
HCV-TR	3b		------A-G----------F----C---GLEYT-TS--
Z4	4a		EHY--AS-I
GB809-4	4a		EHY--AS-I
Z1	4b		VHY--AS-V
GB116	4c		-E----AV---I----------S----------T--VNY--AS-V
GB215	4c		IHY--AS-V
GB358	4c		VNY--AS-I
DK13	4d		------L----------NY---S-V
CAM600	4e		------AV---I----------T--VNY--AS-I
GB809-2	4e		------AV---I----------GVNY--AS-V
CAMG22	4f		------AV---I----------VHYH-TS-I
CAMG27	4f		VHYH-TS-I
GB549	4g		------AV---I----------QHY--IS-I
GB438	4h		------AV---I----------V---R-------QHY--AS-I
BNL7	4k	30	--I-F----------INY--VS-I
BNL8	4k	32	--I----------INY--TS-I
BNL9	4k	34	--I----------INYH-TS-I
BNL10	4k	36	--I------I---X-----X----------TNY--VS-I


BNL11
BNL12
BE95
BE100
HK2
VN4
VN12
FR1
NE98
4k 38
41 40
5a
5a
6a
7c 44
7d 48
9a 42
10a 52
--I-----X----------TNY--VS-I
--I----------I----QHY--VS-I
----------I----------VPY--AS-I
VPY--AS-I
-----AI---I----------T----LTYG--S--
-----XXI--X----X--XX-X--X----T----AHYT-KS--
-X----AI---I----------X----T----LNYA-KS--
-----AI----------T----I--K-AS-I
--I-F----------F--LT-TAGLEY--AS--

Isolate	Type	SEQ ID	V1 201	V2	V3	V4 250
			YHVTNDCPNSSIVYEAADAILHTPGCVPCVREGNASRCWVAMTPTVATRD			
HCV-1	1a		-------S----------M-M----------S-F----------L---L-A-N			
HCV-J	1b		------S----I--MDGM-M-Y----------D-HL---M-L---L-VKX			
BNL1	1d	4	--L----S----I--MSGM---A----------N-S---MXL---L-VK-			
BNL2	1d	8	------S-G------K-I----------X---I----I----PL---L-A-I			
FR2	1f	12	-M----T-D--TWQLQA-V--V----------EKV--T----IPVS-N--VQQ			
HC-J6	2a		-YA----S-N--TWQLT--V--L----------ENDNGTLH--IQV--N--VKH			
HC-J8	2b		-M----S----WQLEG-V----------EQI----------PVS-N--I-Q			
CH610	2c		-MP---S----WQLEG-V----------E-TA-V----------PVA-NL-ISQ			
S83	2c		-M----Q----WQLR--V--V----------EEK--I----IPVS-NI-VSQ			
NE92	2d		-MA---S-N---WQLX--V--V----------ENSSGRFH--IPIS-NI-VSK			
BNL3	2e	16	-MA---A-D---WQLR--V--V----------E-S--RTF--T-VS-N--VSR			
FR4	2f	18	-MA----S-N--IWQMQG-V--V----------ELQ--K----IPV--N--VNQ			
BNL4	2g	20	-M----S----WQLK--V--V----------E-HQ-Q----IPV--N--VSQ			
BNL5	2h	24	-M----S----WQLEE-V--V----------EWKD-T----IPV--NI-VSQ			
BNL6	2i	26	-M----S-X--TWQLXX-V--V----------E----S----IPV--X-XVSR			
FR13	2k	76	-VL---S----------D-V----------I---QD--T-T--TPV----VKY			
BR36	3a		-VL---S-G----------E-V---L----------TT--Q-S--TTVST---V-T			
HCVTR	3b		--I----------DHH---L----------MT--T----TPV----VAH			
Z4	4a		--I----------V---TDHH---L----------A--V----TPV----AVS			
GB809-4	4a		----------T----TEHH-M-L----------TE-T----PL----APY			
Z1	4b		--I----------DYH---L---L---V--Q----------L----APY			
GB116	4c		----------DHH---L---L---V--Q----------LS---APY			
GB215	4c		--I----------TEHH---L---L---V--Q----------L----APY			
GB358	4c		----------TDYH---L----------K-T--SL----AQH			
DK13	4d		--I------A----TENH---L----------T--Q----------L----SPY			
CAM600	4e		--I------A----TDNH---L----------KT--Q----------L----SPY			
GB809-2	4e		--L----------F--VHH---L----------T--Q----------L---L-APY			
CAMG22	4f		--I----------F--EHH---L----------T--Q----------I-L---L-APH			
CAMG27	4f		----------DHH-M-L----------T--T----------PL----APY			
GB549	4g		-Y----------DHH-M-L----------T--V----------IPL----VPY			
GB438	4h		----------DHH---L---L----------Q----------L----APY			
BNL7	4k	30	--I----------DHH---L---L----------T--Q----------L----APY			
BNL8	4k	32	----------DHH---L---L----------V--Q-S----L---I-APY			
BNL9	4k	34	----------DHH--AL---L----------V--Q----------L----APY			
BNL10	4k	36	----------F--DHH---L---L----------K---H----------L----APY			
BNL11	4k	38				

```
BNL12   41   40   ----------------SDHH---L-------KT--T-------L-----API
GB724   4x        --I---------V---TDHH---L--------T--V-----TPV-----AVS
BE95    5a        -----------------DNL---A-------MT--V------QI---LSAPS
BE100   5a        -----------------D-L---A--------KD-V------QI---LSAPS
HK2     6a        --L-----------L--DAM---L---L----VDDR-T--H-V---L-IPN
VN4     7c   44   --L--------------ETL---L-------KXX-Q------QAS--L-VPN
VN12    7d   48   --L--------------NGM---L-------KT--LTK--LSAS--L-VQN
FR1     9a   42   --L-----S-N---F--ETM---L------IKA--E----LPVS--L-VPN
NE98    10a  52   -M-----S-G------G-I---L---------S--T----IPVSX---VKS
```

Isolate	Type	SEQ ID	V4 251	V5 300
HCV-1	1a		GKLPATQLRRHIDLLVGSATLCSALYVGDLCGSVFLVGQLFTFSPRRHWT	
HCV-J	1b		SSI-T-TI---V-----A-A---M-----------S----------YE-	
BNL1	1d	4	ASV-TXAI---V-----XX-F---M--X----------A----------M-H-	
BNL2	1d	8	ANV-TAAI---V-----T-AFR--M-----------------------LYH-	
FR2	1f	12	ANA-IDEV---V-----A-VF---M-I-----G----------TS-----	
HC-J6	2a		PGALTQG--T---MV-M----------G-M-AA-M-IV--QH--F	
HC-J8	2b		RGALTRS--T-V-MI-MA--A----------V--A-MILS-A-MV--Q--NF	
CH610	2c		PGTLTKG--A-V-VI-M----------V--ALMIAA-AVIA--Q--TF	
S83	2c		PGALTKG--A---II-M---V----------V--ALM-AA-VVVV--QH-TF	
NE92	2d		PGALTKG--T---TIIA---F-----I-----A-M-AS-V-II--QH-KF	
BNL3	2e	16	PGALTKG--AR--AV-M----------V--A-MIAA-A-IVA-K--YF	
FR4	2f	18	PGALTRG--A---TI-M----------I-----A-MIAA-VAVV--QY-TF	
BNL4	2g	20	PGALTRG--T---TI-MV---------I--V--A-MIAA-VVIV--QH-NF	
BNL5	2h	24	PGALTRG--T---TI-A---V----------F--A-M--S-F-MI--QH-IF	
BNL6	2i	26	PGAXTKG--T---II-A---F------	
FR13	2k	76	PGALTEG--S---TI-A---F-----I--V--AIMIAA-VVIV--EH-HF	
BR36	3a		VGATTASI-S-V-----A--M----------M--A-----A---R----Q-	
HCVTR	3b		LGVTTASI-T-V-M---ARQ----------AF-A-----A---R----T-	
Z4	4a		PGA-LESF---V--M--A----------GA--M--MI--R----------	
GB809-4	4a		MDA-LESF---V--M--A--V--V----------GA-----M---Q-----	
Z1	4b		PNA-LESM---V--M--A--M---F-I-----G----------D-R-----	
GB116	4c		VGA-LES--S-V--M--A--V-----I-----G-----M-S-Q-------	
GB215	4c		IGA-VESF---V-MM--A--V-----I-----G-----M-S-R-------	
GB358	4c		IGA-LES--S-V--M--A--A-----I-----G-----M-S-Q-------	
DK13	4d		LNA-LES---V--M--G----------I--V--G----------Q-----	
CAM600	4e		AGA-LEP---V--M--A--M-----I-----GL-----M---Q-------	
GB809-2	4e		VGA-LEP---V--M--A--V----------GL-----M---Q-------	
CAMG22	4f		LGA-LESM---V--M--T----------GI--A--M---R--L-----	
CAMG27	4f		IGA-LESM---V--M--T----------I-----GI-----M-N-R--L---	
GB549	4g		VGA-LESM---V--M--A--V-----I---G-----M---R-------	
GB438	4h		LGA-I-SV-Q-V--M--A----------I--H--G---A--MVS-Q-----	
BNL7	4k	30	IGA-LES--S-V--M--A--V-----I--X-XGL-----M-S-R------	
BNL8	4k	32	IGA-LES--S-V--M--A--V-----I-----GL-----M-S-R------	
BNL9	4k	34	IGA-LES--S-V--M--A--V-----I-----GA-----M-S-R------	
BNL10	4k	36	TAA-LES--S-V--M--A--V-----I-X---GL-----M-SXQ------	
BNL11	4k	38	IGA-LES--S-V-VM--A--V-----I-----GL-----M-S-R------	

```
BNL12    LSA-LMSV---V--M--A---S----------GA-----M---Q------
GB724    VDA-LESF---V--M--A----------V----GA-----M---Q------
BE95     LGAVTAP---AV-Y-A-G-A----------A--AL-----M--YR--Q-A-
BE100    FGAVTAP---AV-Y---G-A----------A--AL-----M--YR--Q-A-
HK2      AST---GF---V---A-A-VV--S--I-----L--A-----Q------
VN4      AST-V-GF-K-V-IM--A-AF---M-------GL----------LR--M-QV
VN12     ASVSIRGV-E-V-----A-AF---M-------GL----------R--MYEI
FR1      SSV-IHGF---V-----A-AF---M-I-----II----------R-KY-QV
NE98     PCAATAS--T-V-MM-XA----------AL--X--G-SWRH-Q---

          41  40
4z        4z
5a        5a
5a        5a
6a        6a
7c        44
7d        48
9a        42
10a       52
```

Isolate	Type	SEQ ID	V5 301	319
			TQGCNCSIYPGHITGHRMA	
HCV-1	1a			
HCV-J	1b		V-D----------VS----	
BNL1	1d	4	--E------	
BNL2	1d	8	--E------	
FR2	1f	12	V-D-------S-------XXX	
HC-J6	2a		V-D----------T-------	
HC-J8	2b		--E------Q----------	
CH610	2c		V-E----------------X	
S83	2c		V-E---------R-------	
NE92	2d		V-D----------------	
BNL3	2e	16	V-E----------------	
FR4	2f	18	V-E----------------X	
BNL4	2g	20	S-D------	
BNL5	2h	24	V-D------	
FR13	2k	76	V-D-------------P--X	
BR36	3a		V-T-----L-----LS----	
HCVTR	3b		V-T----------VS---	
24	4a		--E------T----------	
GB809-4	4a		--D------T----------	
21	4b		--D----------VS-----	
GB116	4c		--D--------A--V-----	
GB215	4c		--D----------A------G--	
GB358	4c		--D--------A--V------	
DK13	4d		--D------T----------	
CAM600	4e		--D------T----------	
GB809	4e		--D------A----------	
CAMG22	4f		--E----T----------	
CAMG27	4f		--E----------------	
GB549	4g		--D------D----------	
GB438	4h		--D--------V--------	
BNL7	4k	30	--D------	
BNL8	4k	32	A-D------	
BNL9	4k	34	--D------	
BNL10	4k	36	--D------	
BNL11	4k	38	--E------	
BNL12	41	40	V-D------	

```
GB724   4x  --D-------T---------
BE95    5a  V-N-------S--V------
BE100   5a  V-D-------S--V---Q--
HK2     6a  V-D-------T--V------
VN4     7c  V-E-------T---------
VN12    7d  A-D-------A---------
FR1     9a  --D----XNX--V------
NE98    10a V-D-----
```

Figure 5. NS5B nucleotide alignment

```
Isolate  Type SEQ ID   7932                                                7981
                       CTCCACAGTCACTGAGAGCGACATCCGTACGGAGGAGGCAATCTACCAAT
HCV-1    1a            ---A--G----------AT----------T-----AT-----T-------
HCV-J    1b            ---A--G----------AT----------T-----AT-----T-------
BE90     1b            N--A----------C----A----------GTT-----T-----T-------
BNL1     1d       53   ---G-----T----------AT----------GTC-----AT-----A-------
BNL2     1d       55   ---G-----T----------A----------C----RAT----------T----
FR17     1d       57   ---G-----T----------A----------T---GTC-----AT----------G-
CAM1078  1e       61   ---A--G----------AGCT--T-----A--A-----T-C--A-------
FR2      1f       63   N--A----------T--T----------A-----T-C----------
FR16     1g       67   NNNNNNN----------T--T----------GTC-----RT-----T-------
HC-J6    2a            ---A--C----------A----------A-G--T----------T-C--A--T-GGG
HC-J8    2b            ---A--C-----G-----G-----AA-A--A--A--AT-C--A--T--GG
BNL3     2e       69   ---G-----A-----A--T--AA-N--T----------T-C--A-----GG
FR4      2f       71   ---A--C-----A-----G--T--AA-A--T----------T-C--A----TGG
BNL5     2h       73   ---A----------G-G-----A----------A-G--C----------T-C--T----TTG
FR13     2k       77   A----------A--A----------A-AGTT--A---T-CG-T--T-TG-
FR18     21       79   ---A----------G----------G-----A-G-AT----------T-C--A-T--TGG
T1       3a            ---A--T----------ACAG----------A-GGT---A-----AG--A-------
T9       3b            ---T--T----------ACAT----------A-G----------AG--A-------
PAK64    3g       81   ---T--T----------ACAG--T---A-GGTA--A--A-A---A-------
```

Isolate	Type	SEQ ID	7932	7981
GB48	4c		------T--A--C--A-AG-------A-GGTC---------AGG-----T--G-	
GB116	4c		------T--A--C--A-AG-------A-GGTC---------AGG-A--T--G-	
GB215	4c		------T--A--C--A-AA-------A-GGTC---------AGG-A--T--G-	
GB358	4c		------T--A--C--A-AG-------A-GGTC---------AGG-A--T--G-	
GB809	4e		------T--G---------A-------A-GGTC---------AGG-G--T--G-	
GB549	4g		------G--G--C--A--G--T---A-G--C---------A-AG---------G-	
BNL8	4k	83	------T--A--C--A-AG-------A-GC-C---------A-AGG-----T--G-	
BNL12	4l	85	------G--G---------A-AG-------A-GGTC---------A-AG-----T--G-	
EG81	4m	87	------C--A--C--A--G---------A-GGTC---------AGG-----T----	
CHR18	5a		---G--C--T--C--ACAT---------AATG--T--A---T-T--T--------	
VN13	7a	89	---A---------A---C-----TG---AG---------C-T-AC---------G-	
VN4	7c	91	---G---------C---C---------RC--C---C--C-AC----------	
VN12	7d	93	---T-C-------G---C-T---------C--T--AC-C-AC-----T--G-	
FR1	9a	95	A---------G-G-C---------A--C--A--ACNA-AC--T-----TG-	
NE98	10a	97	-----T---------CAG---------A-GGTA--ACTTT-C----TT--GG	
FR14	11a	99	---T--C-------A-------G-----A--G--A--A--AT-C-----T-TG-	
FR15	11a	101	-----T---------A-------A--G--A--A--AT-C-----YYTGG	
FR19	11a	105	---T--T-------A-------G--T--A--A--A-------AT-C--Y--T-TGG	

Isolate	Type	SEQ ID	7982 GTTGTGACCTCGACCCCCAAGCCCGCGTGGCCATCAAGTCCCTCACCGAG 8031
HCV-1	1a		--------T-G-C----G------A-GCA-------A-G---G------A---
HCV-J	1b		--------T-G-C----G------A-GCA-------A-G---G------A---
BE90	1b		--------T-G-C----G-G---A-ACA-------A------G------A---
BNL1	1d	53	--------T-G-C----G-G--T---AA-------A------G---------
BNL2	1d	55	--------T-G-C----YG-G-------AA-------A------G---------
FR17	1d	57	-C------T-G-C----G-G-------AA-------A------G---------
CAM1078	1e	61	-C--------GC-----G----A--T--A-----------TT-G--T--A
FR2	1f	63	-C------T--A-----G-G--T---AA-------A-G---------A---
FR16	1g	67	-C---------G-C----G-G--T------A------A------G------T---
HC-J6	2a		C----TC-T-GCC-GAGG-G----A-ACT------AC-C--A--G--T---
HC-J8	2b		C----TCT--GCCT-AAG------A-AACT-T---AC-C--G------T---
BNL3	2e	69	C----TC-T-ACC-GAG--G---A-AACT-------AC-C--AT-G--T---
FR4	2f	71	CC--CTC-T-ACC-GAG--G------GACT-------AC-T--AT-A--T---
BNL5	2h	73	CC--CTC-T-ACC-GAG--------AACT-------AC-T--AT-G--T---
FR13	2k	77	CC---TCA--TCC-GAGG-G------A-CT-------AC-C--A--A--T---
FR18	21	79	CC--CTCGT-GCC-GAGG-G------GACT-T---AC-T--G------T---
T1	3a		-C---A----T--A--GG-G---A-GAGA-TG---TCC----------G---
T9	3b		-C---------T--G--AG-G--T--GAA---G------GCG-T------A---
PAK64	3g	81	----------T--G--GG-G--TA-ACG---A------A--G------G--A

Isolate	Type	SEQ ID	7982	8031
GB48	4c		----------G--G---G----------AA---A--T-CCG----A--A---	
GB116	4c		----------G--G---G-G--------AGA--A--T-CCG----A--A---	
GB215	4c		----------G--G---G----------AA--TA--T-CCG----A--A---	
GB358	4c		----------G--G---G-G--------AA---A--T-CTG----A--A---	
GB809	4e		-------T--G--G---G-G--------AA--TA--AGCCG----------G---	
GB549	4g		-C--C----G--G---G----------AA--TG--ATCCG----A--G--A	
BNL8	4k	83	----------G--G---G----T---AA--TT--T-CCG----------A--A	
BNL12	41	85	----------G--R---G----------AAA--A--ATCCG----------A---	
EG81	4m	87	----------T--G--G--AG-G----------AA---A--ATCCG----------G---	
CHR18	5a		CA-TGT--T-GC-G--TG-G--G--T----------A--ACG----A----------C-A	
VN13	7a	89	-C---A-GT-G--G---GC---A--GACA----------CA--G--T--T--C	
VN4	7c	91	-C--CC-A--T-------GGTG--A---AA---T--T-CA--T--G--T---	
VN12	7d	93	-C--CC-AT-A--T--GGT---A--GAAA----------T-CA--T--T--T---	
FR1	9a	95	CC--CC-G----------AG-G-----GAAA----------A----------T----T---	
NE98	10a	97	CC----------A-GGA-G-G--TA-GAG--TG--A-CT--A----------G---	
FR14	11a	99	C----C-AT-GCCTGAAG-G-----GAAA----------T--A--G--G--A---	
FR15	11a	101	C----C-AT-GCC-GAAG-G-----GAA----------T--A--A--G--A---	
FR19	11a	103	C----C-AT-GCC-GAAG-G-----GAA----------A--A--G--A---	

Isolate	Type	SEQ ID	8032 AGGCTTTATGTTGGGGGCCCTCTTACCAATTCAAGGGGGGAGAACTGCGG	8081
HCV-1	1a		C---------C-----T--C--G--T-----G-A----C----------	
HCV-J	1b		C---------C-----T--C--G--T-----A----C----------	
BE90	1b		C---------A-C-----T--C--G--T-----A----C----------	
BNL1	1d	53	C----G--CA-C---------Y--A---------AA--AC----------	
BNL2	1d	55	C----G--C--C---------C--A---------A----C----------	
FR17	1d	57	C----G---A-C---------T--C--A---------C----AA---C--	
CAM1078	1e	61	-----G--C--C---------G--CT-G---------AA----------	
FR2	1f	63	---------A-C---------T--C--G--A--C----AA---C----	
FR16	1g	67	C----A-----C---------C-----C----AA--AC----------	
HC-J6	2a		--A-----C--G--A--G--CA-GTT---CAGC-A---CC---C----	
HC-J8	2b		--A-----C--A--A--G--CA-G--A--CAGC-AA---C-ATC----	
BNL3	2e	69	--A--C--C--A--A--G--CA-G-TG--CAGC-AA---C-ATC----	
FR4	2f	71	--A-----C--G--A--G--CA-G-TG--CAGC-AA---C--TC----	
BNL5	2h	73	--A-----C--A--A--G--CA-G-TG--CAGC-A---AC--TC----	
FR13	2k	77	-----G--C--G--A--T--CA-GCAG--CAGC-A----C-ATC----	
FR18	21	79	--A--C--CA-A--C--G--GA-G-TG--CAGC-AA--CC--TC---T--	
T1	3a		C---------CTGC---------A-GTT---CAGC-A----CCC-A--T--	
T9	3b		C----G--CA-C--A--T--CA-GTA---CAGT-A----CTCC-G----	
PAK64	3g	81	C---------C-----A--T--CA-GTT---CAGC-A----CTC--A----	

Isolate	Type	SEQ ID	8032	8081
GB48	4c		--A--C--C--G--C--T--CA-GCAT--CAGC-A--A--CCTG----	
GB116	4c		--A--C--C--G--C--T--CA-GCAT--CAGC-----A--CCTG----	
GB215	4c		--A--C-----G--C--T--CA-GCAT--AGC-AA--A--CCTG----	
GB358	4c		--A--C-----G--C--T--CA-GCAT--CAGC-A--A--CCTG--T--	
GB809	4e		--A--C--C--G--C-----CA-GCAT--CAGC-A--A--CCTT----	
GB549	4g		--A--C--C--G--C--T--CA-GTA---C--C-A-----CCTA----	
BNL8	4k	83	--A--C--C--G--C-----CA-GCA---CAGC-A--A--CCTT--T--	
BNL12	4l	85	--R--C--CT-G--C-----CA-GTAT--CAGC-AA-----CT-----	
EG81	4m	87	--A--C-----G--C--T--CA-GTTT--CAGC-A--A--CCTA--T--	
CHR18	5a		C-C--G--CTG---A-----CA-GTAT--CAGC-A----C-AC-A--T--	
VN13	7a	89	C-AT-G--CTNC--T--T--CA-GTNT--C--T-AA--TC--GCA--T--	
VN4	7c	91	C----G--CTGC--W--G--CA-G-TG--C--CC-T--TC-ATCA--T--	
VN12	7d	93	C----G--CTGC--C-----CA-GTA---C--TC-A--TC--TCA--T--	
FR1	9a	95	-----C-----C-------A-GTA---C----A---CC-ACT---T--	
NE98	10a	97	C-------CTG---T--T---A-GTT---CAGC-A---AC-AC-----	
FR14	11a	99	--A--A--C--G--C-----GA-GGAA--CAGC-A---CC--GCT----	
FR15	11a	101	--A--A--C--G--C-----GA-GGAA--CAGC-AA--CC--GC-----	
FR19	11a	105	--A--A--C--G--C-----GA-GGAA--CAGC-A---CC--GC-----	

Isolate	Type	SEQ ID	8082 CTATCGCAGGTGCCGCGCGAGCGGCGTACTGACAACTAGCTGTGGTAACA 8131
HCV-1	1a		
HCV-J	1b		T-----C----------A--T-----G-----G-------C--C----
BE90	1b		-----C-A---------A---------G---C-------C-------T-
BNL1	1d	53	---C--TC---------C---------T---C---C--C----
BNL2	1d	55	-----TC----T-----T---------G-----C---C--C----
FR17	1d	57	---C--TC---------C---------G-----T---C-------T-
CAM1078	1e	61	---------A-------T--C---------CT--------C--C----
FR2	1f	63	---C---C-A-------T--A---------C-----G-------C----
FR16	1g	67	-------C---------T---------T--G-------T--------
HC-J6	2a		G--CA-GC-T---------C-------G--G--T--C-------ATG--G----
HC-J8	2b		---CA-GC-T---------A-------T--TT-C--C--C--ATG--G--T-
BNL3	2e	69	A--CA-GCAT---------C-------A--G--C--C--C--TATG--G--T-
FR4	2f	71	A--CA-GC-T---------T-------A--G--C--C--C--TATG--G----
BNL5	2h	73	T--CA-AC-T---------C-------A--G--C--C--C--ATG--G--T-
FR13	2k	77	A--CA-GC-C---------C-------G--G--C--C-------ATG--G--T-
FR18	2l	79	A--CA-GC-T--T------C-------G--GT-C--C-------ATG--C--T-
T1	3a		T-------C---------T--C--T--A--C---C-T--C----TC--C----
T9	3b		-------C-C---------C-------------CT--C-T--C---TC--C--T-
PAK64	3g	81	A-----C-T---------T--T-------T--T---C-C-------AC----T-

Isolate	Type	SEQ ID	8082	8131
GB48	4c		G---------A--T-----A---------CTAC--C--C---TC--G----	
GB116	4c		G---------A-----T---------CTAC--C--C---TC--G----	
GB215	4c		G---------A---------A---------CTAC--C--C---TC--G----	
GB358	4c		G---------A---------A---------CTAC--C--C---TC--G----	
GB809	4e		G-----T--A---------A---------TAC--C--C---TC--G----	
GB549	4g		GC-A--G---------A-----G--CTAC--C--C---TC--G----	
BNL8	4k	83	G-----G--A---------A---------CTAC--G--C---TC--A----	
BNL12	41	85	G-----G---------A--GTAC--C--A--T-TC--G----	
EG81	4m	87	---C---------A---------CTAC--C--C---TC--A----	
CHR18	5a		T-----T--A---------C---------CT-C--C---TATG--C----	
VN13	7a	89	A--C--T---------A-G--C--T---------CT---C--C-T-CTG-CC--T-	
VN4	7c	91	A--C--T---------A--C--T---G--C--C--G---TG--C--T-	
VN12	7d	93	G--C---------G--T--T--T--CT-C--C--A---TG--C----	
FR1	9a	95	TC-A---C-A---A--A---A---CC-C--A--ATG------	
NE98	10a	97	T--C---C-C---T--T--T--G--G--AC-C--C---TC--G----	
FR14	11a	99	A---A-GC-T---------A---G---T-C--C--A---TG--G----	
FR15	11a	101	A---A-GC-T---------A---G---T-C--C--A---TG--G----	
FR19	11a	105	A--CA-GC-T---------A---G---T-C--C--A---TG--G----	

Isolate	Type	SEQ ID	8132 CCCTCACTTGCTACATCAAGGCCCGGGCAGCCTGTCGAGCCGCAGGGCTC 8181
HCV-1	1a		
HCV-J	1b		------A--T---T-G-------ACT--G----------T---AA----
BE90	1b		----T--A--T---C-A-------TCT----------T--GAA----
BNL1	1d	53	----G--A-------T-G--A---A-A--G----------T---AA----
BNL2	1d	55	-------A-------T-G--A---A----G----------T---AA----
FR17	1d	57	-------A--T---T-G--A---A----G----------T--GAA----
CAM1078	1e	61	-------C------T----------TA----------A----T--CAA----
FR2	1f	63	-------C--T------A----------A----------T--GAA----
FR16	1g	67	-------A-------C-G--A---GCC--G----------T---AA----
HC-J6	2a		--A----A-------TG-G--A---TTA--G----------AAG--T----------A-A
HC-J8	2b		--A-G--A--T----------A----TT----------G---AAG--T----------A--
BNL3	2e	69	--A----G----------TA--G--T---AA---A----AA-A
FR4	2f	71	--A----G--T--TG-G--A----TC----------T---AA---T--G--CA-T
BNL5	2h	73	--A----A----TG-G----------ATTA--T----------CAA---T----CA--
FR13	2k	77	-T-----A-------T-G----------A----G----------CA-G----G--CA-T
FR18	2l	79	--A----G----TG-G--A---AT----------T--CA----T--C---A-T
T1	3a		-AA-------T----------ACA--G--TGCGAAG----------C---
T9	3b		-AA-A--C--T----------ACT----------A-CA-G--T--G--T---
PAK64	3g	81	-AA----C----------A-A--G--TGC----------T--G--C--T

Isolate	Type	SEQ ID	8132	8181
GB48	4c		-A--G--G------C----A---TCA--C--TATCAA---G--G----G	
GB116	4c		-A--G--G------TC----A---TCA--C--TATCA----G--G----G	
GB215	4c		-A--G--G------TC----A---TCA--C---ATCA-G--GT-----G	
GB358	4c		-A--G--G------C----A---TCA--C--TATCA----G--G----G	
GB809	4e		-AA-G--G------C-T-------TCA--------ATCA-G--T--G----A	
GB549	4g		-TG-A--G--T--TC----------GTT--G--TAC-A-G--------T--G	
BNL8	4k		-A--G--G------C----A---TCA--T--TAT-A----G--------G	
BNL12	41		-AG-G--C------TC-T-------ACC--------TACCA-G--T----C--A	
EG81	4m		-A--G--C------C----------AC---C--TACCA----G--C--C--G	
CHR18	5a		--A-G--G---------T------TTA--CT-----A---------AA----	
VN13	7a	89	-T--G----------T-G--A--T-A---G--A--CA----T--C----G	
VN4	7c	91	----A--A------TT-G--A--A-AA------G---A-G--A---AA----	
VN12	7d	93	--A-G--A------C-G------T-A---G--T---A-G--A--RAA-----	
FR1	9a	95	--A----A----T-C-G------AACC--C--T--C------A--C--CT-T	
NE98	10a	97	-AA----C--T----------A--AAA-------TACCAA---T--C--AA-T	
FR14	11a	99	--A-G------T----------A--TAAA--G--T---AA-------T--CA-T	
FR15	11a	101	--A-G------T----------AAR-----T---AA---Y--T--CA-T	
FR19	11a	105	--A-G------T----------A---AA---G--T---AA-------T--CA-T	

Isolate	Type	SEQ ID	8182 CAGGACTGCACCATGCTCGTGTGTGGCGACGACTTAGTCGTTATCTGTGA 8231
HCV-1	1a		
HCV-J	1b		----------G----------AAC--A----------C-T----------
BE90	1b		----------G----------C--G----------C-T----------
BNL1	1d	53	-G--------------C--G--T---C-T----------
BNL2	1d	55	----------G----------C--A----------C-T----------
FR17	1d	57	----------A----------C--A----------C-T----------
CAM1078	1e	61	----------C----------C-G----G----C--
FR2	1f	63	----T---T-A----------C-T----------
FR16	1g	67	-G---A----------C-C----------
HC-J6	2a		ATT-CGCC---A----G--A--C----T----G--T--C----CA--
HC-J8	2b	69	GT----CCTGTT---T-G----------A----------C-G----C----CA--
BNL3	2e	71	GT--C-CC----------G----C----T---C----T--C----CA--
FR4	2f	73	GTT-C-CC---G----C----T---C-G--T--C----CA--
BNL5	2h	77	GTT-CTCC---G----G--T----------TC-G---A-C----CA--
FR13	2k	79	GTT-CACC---A----G----------C-G----C----CA--
FR18	2l		G-C-C-CC---A---T-G--A----------C-G--G--G--C----CA--
T1	3a		-G-A--CCGGA-T-T--T--C--C--A--T--TC-G----AG-GGC---
T9	3b		A-A---CCAT-TT-C--T--C--C--A----T--G--G--G--G-A-C---
PAK64	3g	81	--A---CCAT-AT-C--T--C--C--A--T--T--G--G--AG-GGC---

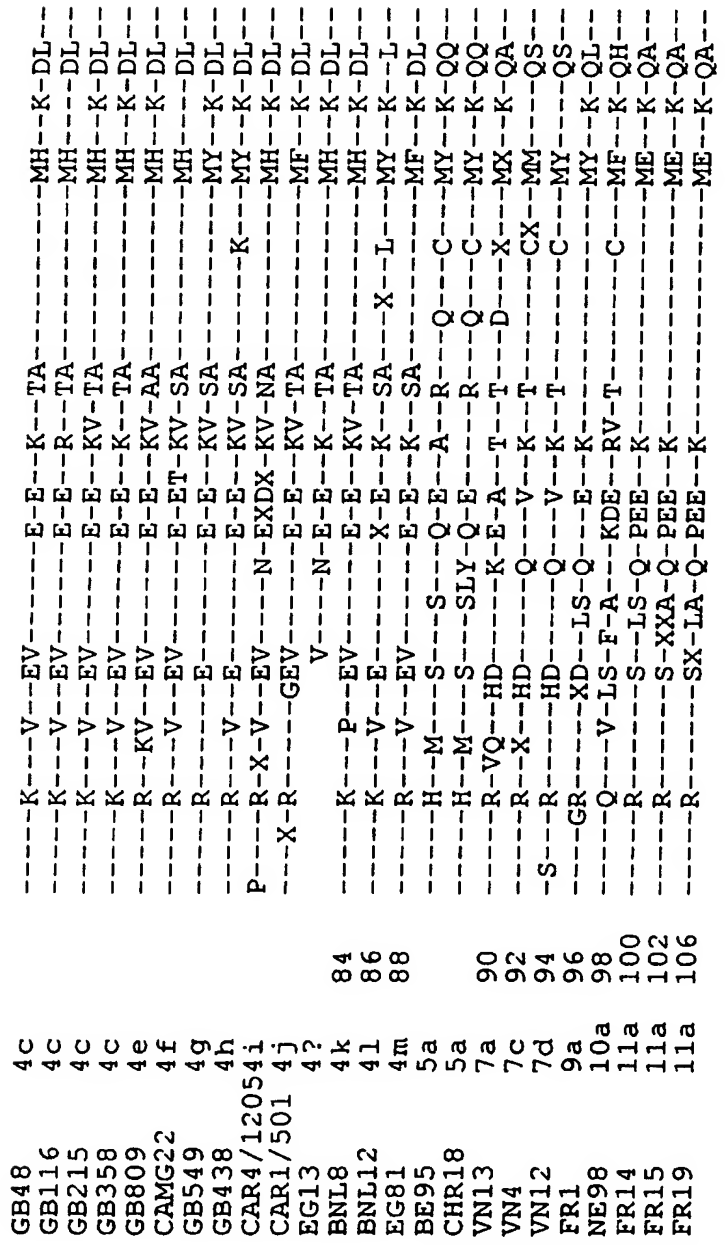
Isolate	Type	SEQ ID	8182	8231
GB48	4c		AGA----------T-G--C-----T--T--C-G--T--C---GC---	
GB116	4c		AGA----------T-G--C-----T--T--C-G------C--TGC---	
GB215	4c		AGA----------T------G--C-A---T-------C-G------C--TGCC--	
GB358	4c		AGA----------T-G--C-----T--T--C-G------C---GC---	
GB809	4e		A-----T----------G--T--C--T----------C----------G---GCC--	
GB549	4g		A-A-GT---G----------G--T-------A---------------C------C--	
BNL8	4k	83	AGA----------G--T--C--T----------G------C---GC---	
BNL12	41	85	A-A----------G--C--C--T----------G------C---GCC--	
EG81	4m	87	A-----T------A----G--T--C--G----------C-G------C---GCA--	
CHR18	5a	89	----------GC-C--G----------T--T--TC-T--G-CC--T--C--	
VN13	7a	91	A----------TTGA-----T-G--C--C--A------C-T----------T-CG--	
VN4	7c	93	A-AA-----ATGA-----T-A--C--C--A------TC-------------GCG--	
VN12	7d	95	A-AA-----TTGA-----T-G--C--C--A------C-------------TGC---	
FR1	9a	97	ACA--T-ATGA-----T-G--C--C--A------T--G--T--CG-AAC---	
NE98	10a	99	A-AA-TCCAT-AT-C--T--C--C--A--T------G------G--TGC---	
FR14	11a	101	GTA---CCGGTG----------C--T----------C----G--C----CA--	
FR15	11a	105	GTT---CCGGTG----------C----------C----G--C----CA--	
FR19	11a		GTT---CCAGTG----------C----------C----G--C----CA--	

Isolate	Type	SEQ ID	8232 AAGCGCGGGGGGTCCAGGAGGACGGCGGCGAGCCTGAGAGCC 8271
HCV-1	1a		
HCV-J	1b		G--T-----AAC----------T------GC---AC------
BE90	1b		----------AAC---A----------------AC---T-
BNL1	1d	53	G--T------A---G----A----------A---AC---T
BNL2	1d	55	G----------A---G----------A---AC---T-
FR17	1d	57	G--T--R--A---G----------T----------A---AC---T-
CAM1078	1e	61	G--T-TA---AC----------T----------C------
FR2	1f	63	G-T--A-----N-----TC--T---
FR16	1g	67	G--T----------T--T--A------
HC-J6	2a		G---CA-----AC-G----------A-CG--A----------
HC-J8	2b		G---CAA--TAA-G----------A-CGA-A---------T
BNL3	2e	69	G--TCA---A---G----------ACCG--A------
FR4	2f	71	G--TCA-----CTG----------A-CGA-A---------T-
BNL5	2h	73	G--TCA---AAC-G----------T-A-CG--A---------T-
FR13	2k	77	G--TCA----ACTG--AG-----A-AAC-A---------C-T
FR18	21	79	G--TCA----AC-G----------A-CGA-AT---------T-
T1	3a		G----AT--C---G-T-----TAGA--AGC----------
T9	3b		----TGC--C---G----------AGA--AGCT---C-----
PAK64	3g	81	G--TTGC-KC--TG-T-----G-ATAG-GCAGC

Isolate	Type	SEQ ID	8232	8271
GB48	4c		G---AT--C--AG---------AAACGACC---CG------	
GB116	4c		-----AT--C--AG---------AAACGAGC---CG------	
GB215	4c		G---AT--C--AG---------AAACGAGC---CG---T-	
GB358	4c		G---AT--C--TG---------AAACGAGC---CG------	
GB809	4e		G---GT--C--TG---------AAACGANC---CG---T	
GB549	4g		G---GC--C--AG---------T--AAGAGC---CC-----	
BNL8	4k	83	G---AT--C--AG---------TAACCGAGC---CCN----	
BNL12	41	85	G---A---C--AG---------TT-CCAACC---CC-----	
EG81	4m	87	G---AT--C--GG-C---------CGCCGAGC---CCA---T	
CHR18	5a		G---CA----ACG--C---------TAAA-------------	
VN13	7a	89	G--TTT----------TC---------A-TAGTGCA---C----T	
VN4	7c	91	G--T-GA--A---TCT---------T-TT-ACGC---C----A	
VN12	7d	93	G---GA--A----CT---------T--C-G-GC---C---T-	
FR1	9a	95	G--T--T--A---A-C---------TATC--T-A---C-----	
NE98	10a	97	G--T--A---A--G-T---------AA-AGCGC-T-------T	
FR14	11a	99	----AA----------GG---------CA-CG-GA---AC---T-	
FR15	11a	101	G---AA----------AG---------CA-CGAGA---AC	
FR19	11a	105	----AA----------GG---------CAACGAGA---AC--NT-	

Figure 6. NS5B amino acid alignment

Isolate	Type	SEQ ID	2645	2694
HCV-1	1a		STVTESDIRTEEAIYQCCDLDPQARVAIKSLTERLYVGGPLTNSRGENCG	
HCV-J	1b		-----N------S-------A-E--Q--R---------------K-Q---	
2TY4	1c		-------H-D--A---N--------------K-----	
BNL1	1d	54	-----N---V--S-------A-E--K----------I--X----K-Q---	
BNL2	1d	56	-----N-----XS-------AXE--K------------------K-Q---	
FR17	1d	58	-----N---V--S-------A-E--K----------I-------K-Q---	
CAM1078	1e	62	-----A------S-------H-E---------------------K-----	
FR2	1f	64	----------S---------E--K--R-------I-------K-Q---	
FR16	1g	68	XX-------V-XS-------A-E---------------------K-Q---	
HC-J6	2a		-----R------S--RA-S-PEE-HT--H----------MF--K-QT--	
HC-J8	2b		-----R------S---A-S-PQE--TV-H----------M---K-QS--	
ARG8	2c		---S-S-PEE--T--H----------M---K-QS--	
NE92	2d		-----R------S--LA-S-PE---T--H----------ML--K-QT--	
BNL3	2e	70	-----R--X---S---A-S-PE---T--H----------MM--K-QS--	
FR4	2f	72	-----R------S--LA-S-PE---T--H----------MM--K-QS--	
BNL5	2h	74	---A-R------S--LA-S-PE---T--H----------MM--K-QS--	
FR13	2k	78	-----R---V--SV-LS-S-PEE--A--H----------MQ--K-QS--	
FR18	2l	80	-----R---N--S-FLA-S-PEE--TV-H-------I---MM--K-QS--	
BR34	3a		------C---MF--K-AQ--	
BR36	3a		------C---MF--K-AQ--	
BR33	3a		------C---MF--K-AQ--	
T9	3b		-----H------E-------E-E--K--SA------I---MY--K-LQ--	
PAK64	3g	82	-----Q---V--E-------E-E--R----------MF--K-LK--	


GB48 4c ----K--V--EV-----E-E--K--TA-----MH--K-DL--
GB116 4c ----K--V--EV-----E-E--R--TA-----MH----DL--
GB215 4c ----K--V--EV-----E-E--KV-TA-----MH--K-DL--
GB358 4c ----K--V--EV-----E-E--K--TA-----MH--K-DL--
GB809 4e ----R--KV--EV-----E-E--KV-AA-----MH--K-DL--
CAMG22 4f ----R---V--EV-----E-ET-KV-SA-----MH----DL--
GB549 4g ----R-------E-----E-E--KV-SA-----MY--K-DL--
GB438 4h ----R---V--E-----E-E--KV-SA-----K---MY--K-DL--
CAR4/12054 4i P---R-X-V--EV----N-EXDX-KV-NA-----MH--K-DL--
CAR1/501 4j ---X-R-----GEV-----E-E--KV-TA-----MF--K-DL--
EG13 4? V---N-E-E--K--TA-----MH--K-DL--
BNL8 4k 84 ----K---P--EV-----E-E--KV-TA-----MH--K-DL--
BNL12 4l 86 ----K---V--E-----X-E--K--SA---X--L---MY--K--L--
EG81 4m 88 ----R---V--EV-----E-E--K--SA-----MF--K-DL--
BE95 5a ----H--M---S---S---Q-E--A--R---Q---C---MY--K-QQ--
CHR18 5a ----H--M---S---SLY-Q-E-----R---Q---C---MY--K-QQ--
VN13 7a 90 ----R-VQ--HD-----K-E-A--T--T---D---X---MX--K-QA--
VN4 7c 92 ----R--X--HD-----Q---V--K--T-----CX--MM---QS--
VN12 7d 94 -S---R-------HD-----Q---V--K--T-----C---MY---QS--
FR1 9a 96 ----GR-----XD--LS-Q---E--K-----MY--K-QL--
NE98 10a 98 ----Q---V-LS-F-A---KDE--RV-T-----C---MF--K-QH--
FR14 11a 100 ----R-------S---LS-Q-PEE--K-----ME--K-QA--
FR15 11a 102 ----R-------S-XXA-Q-PEE--K-----ME--K-QA--
FR19 11a 106 ----R-------SX-LA-Q-PEE--K-----ME--K-QA--

Isolate	Type	SEQ ID	2695 YRRCRASGVLTTSCGNTLTCYIKARAACRAAGLQDCTMLVCGDDLVVICE 2744
HCV-1	1a		
HCV-J	1b		----------L--T-------K----------N----------
2TY4	1c		----------L--------R---
BNL1	1d	54	----------L----------K-R----------
BNL2	1d	56	----------L----------K----------
FR11	1d	58	----------L----------K----------
CAM1078	1e	62	----------L----------K----------
FR2	1f	64	----------K----S----------
FR16	1g	68	----------L--A----K-RE----------
HC-J6	2a		----------M--I--V--L--K--IIAP----------S-
HC-J8	2b		----------F--M--M-----L--K--IV-PV----------S-
ARG8	2c		----------A----M----------V----------N--IVAP-
NE92	2d		----------F--M--I--V--Q--K--IIAP----------S-
BNL3	2e	70	--H----------M--I----------L--K--IVAP----------S-
FR4	2f	72	----------M--I--V--L--K--IVAP----------S-
BNL5	2h	74	----------M--I--V--L--K--IVAP----------I-S-
FR13	2k	78	----------M----------L--Q----------IVAP----------S-
FR18	2l	80	----------F--M--I--V--M----------IDAP----------S-
BR34	3a		----------P--F--I----------T--A----RNPDF----------VA-
BR36	3a		----------P--F--I----------T--AK---RSPDF----------VA-
BR33	3a		----------P--F--I----------T--AK---RNPDF----------VA-
T9	3b		----------P--F--I----------K--S----K-PSF----------VS-
PAK64	3g	82	----------P--Y--I----------A----------PSF----------VA-

```
GB48      4c  ---------Y---F---------L--S--IK----R-----------A-
GB116     4c  ---------Y---F---------L--S--I-----R-----------A-
GB215     4c  ---------Y---F---------L--S--I--S--R-------Y-----A-
GB358     4c  ---------Y---F---------L--S--I--S--R-----------A-
GB809     4e  ---------Y---F---M-----L--S--I-----K-----------A-
CAMG22    4f  ---------Y---F-----FL--T--TK----K-----------A-
GB549     4g  Q--------Y---F---V---L--V--T-------KG-S---------
GB438     4h  L--------Y---F---V---L--T--T-------K-----------A-
CAR4/120541 I--------Y---F---------L--T--T-------K-----------A-
CAR1/501  4j  Q--------F---F---------L--T--T-------K---------S-
EG13      4?  ---------F---F---------L--T--I-----R---
BNL8      4k  ---------Y---F---------L--S--I-----R-----------A-
BNL12     4l  ---------Y---F---V-----L--T--T-----K-----------A-
EG81      4m  ---------Y---F---------L--T--T-----K-----------A-
BE95      5a  ---------F---M---M-----L-S----R-R---L-----------A---
CHR18     5a  ---------F---M---M-----L-S----K-----L-----------A---
VN13      7a  90-------------ILA-----------L--Q-----K-FD----------S-
VN4       7c  92-------------L---------L--Q-----K-KNYD----------A-
VN12      7d  94---------F---L---M---L--Q-----XK-KNFD----------A-
FR1       9a  96---------P--M---I---FL--T-----FT-YD----------VT-
NE98      10a 98---------P--F---I---------K--TK----IKNPSF-----------A-
FR14      11a 100--------F---L---M---------K---K---IV-PV----------S-
FR15      11a 102--------F---L---M---------X---KX--IV-PV----------S-
FR19      11a 106--------F---L---M---------K---K---IV-PV----------S-
```

Isolate	Type	SEQ ID	2745 SAGVQEDAASLRA 2757
HCV-1	1a		
HCV-J	1b		---T-----A---
BE90	1b		---T--------V
BNL1	1d	54	----E----N---
BNL2	1d	56	----E----N--V
FR17	1d	58	-X--E----N--V
CAM1078	1e	62	-V-T---------
FR2	1f	64	IE-XX--PS
FR16	1g	68	----------
HC-J6	2a		-Q-TE--ERN---
HC-J8	2b		-Q-NE--ERN---
NE92	2d		-Q-TE--ERN---
BNL3	2e	70	-Q--E--DRN-
FR4	2f	72	-Q-AE--ERN--V
BNL5	2h	74	-Q-TE--ERN--V
FR13	2k	78	-Q-TER-ENN--P
FR18	21	80	-Q-TE--ERN--V
BR34	3a		-
BR36	3a		-
BR33	3a		-
T9	3b		-C--E--R-A---
PAK64	3g	82	-CX-D-EDRAALR

GB48	4c		-D--E--KRP-G-
GB116	4c		-D--E--KRA-G-
GB215	4c		-D--E--KRA-GV
GB358	4c		-D--E--KRA-G-
GB809	4e		-G--E--KRX-G-
CAMG22	4f		-D--E--RRA-G-
GB549	4g		-G--E---RA---
GB438	4h		-G--E---RA---
CAR4/12054i			-I-ID--KQA--T
CAR1/501	4j		----E--PXTX-P
BNL8	4k	84	-D--E--NRA-X-
BNL12	4l	86	-E--E--SQP---
EG81	4m	88	-D--D--RRA-Q-
BE95	5a		-Q-TH--E-----
CHR18	5a		-Q-TH--K-----
VN13	7a	90	-L--S--TSA---
VN4	7c	92	-G--S--VDA---
VN12	7d	94	-G--P---GA--V
FR1	9a	96	----N--I-N---
NE98	10a	98	---ID--KSA---
FR14	11a	100	-K--E--QRD--V
FR15	11a	102	-K--E--QRD-
FR19	11a	106	-K--E--QRD--